PATHWAY OPTIMIZATION AND ENGINEERING FOR BIOFUEL PRODUCTION

BY

DAWN ERIKSEN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Chemical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Adviser:

Professor Huimin Zhao

Abstract

Optimizing metabolic pathways is paramount for effective and economical production of biofuels and specialty chemicals. One such significant pathway is the cellobiose utilization pathway, identified as a promising route for efficient biomass utilization. Here we describe the simultaneous optimization of the β-glucosidase (*gh1-1*) and the cellodextrin transporter (*cdt-1*) through directed evolution of the pathway. The improved pathway was assessed based on specific growth rate on cellobiose, with the final mutant exhibiting a 42% increase over the wild-type pathway. Metabolite analysis of the engineered pathway presented a 54% increase in cellobiose consumption (1.68 to 2.82 g cellobiose/(L·h)) and a 74% increase in ethanol productivity (0.59 to 1.03 g ethanol/(L·h)). By simultaneously engineering multiple proteins in the pathway, cellobiose utilization by *S.cerevisiae* was improved. This strategy can be generally applied to other metabolic pathways, provided a selection/screening method is available for the desired phenotype. This improved cellobiose utilization *in vivo* will not only decrease the *in vitro* enzyme load in biomass pretreatment, it will also reduce the diauxic shift in pentose sugar utilization, thus significantly reducing the high economics of biofuel processes.

More than engineering microbes to more efficiently utilize the biomass sugars, constructing and designing pathways for biofuel production is also very significant. We explored the development of a biodiesel production pathway using a heterologously expressed fatty acid synthase coupled with a wax ester synthase. In this reaction, the esterification of a fatty acyl-CoA and fatty alcohol catalyzed by the wax ester synthase produces free fatty acid ethyl esters, otherwise known as biodiesel. Only initial experiments have been completed in this project, including initial enzyme characterization and plasmid construction.

This is dedicated to my parents

Table of Contents

Chapter 1: Sustainable Energy Production

1.1 Sustainability

Dwindling natural resources, including fresh water and petroleum, is causing industries and governments to reconsider design principles of industrial processes. Sustainability has now become a major component of all sectors of industrial, governmental, and even societal agencies. Though the definition of sustainable development has been argued over the years, a standard definition has emerged as ". . . a process of change in which the exploitation of resources, the direction of investments, the orientation of technological development, and institutional change are all in harmony and enhance both current and future potential to meet human needs and aspirations. . . (It is) meeting the needs of the present without compromising the ability of future generations to meet their own needs" (1).

1.1.1 Sustainable Development and the Energy Crisis

Sustainable development has been cited as the most difficult challenge that humanity has ever faced (1). To acquire sustainability, many fundamental issues at local, regional, and global levels must be addressed. Industrial sustainability will require a global vision that holistically considers economic, social, and environmental sustainability, which considers more than just short-term profit margins, instead considering long-term investments (1). Part of the challenge in sustainable development is that if a product is made sustainably, it must also be competitive with the conventional counterparts. It must be cheaper, more durable, less toxic, and easily recyclable. These products need to be derived from renewable resources and contribute minimally to net generation of greenhouse gases (1). The technologies to produce those sustainable products competitively in the 21st century are only just emerging.

At the beginning of the 20th century, many commodity chemicals such as dyes, solvents, and synthetic fibers were made from renewable sources such as trees and agricultural crops. By the late 1960s, many of these bio-based chemical products had been replaced by petroleum derivatives (2). In recent decades, liquid petroleum consumption has increased dramatically for both commodity chemicals and energy. Currently, $24 billion worth of hydrocarbon feedstock is used annually in the chemical industry (1). Though the chemical industry relies on petroleum, the most significant use of petroleum is for energy. In 2011, a global consumption of 87.42 million barrels of petroleum each day (3). Experts predict the energy demand to grow by more than 50% by 2025 (2). The company British Petroleum (BP) has mapped energy consumption since 1986 and the trends energy demand increases at an alarming rate (Figure 1.1) (4). In one single year, from 2010-2012, energy consumption grew by 2.5%, with China alone accounting for 71% of global energy consumption growth (4).

Though the demand for energy is ever increasing, peak oil production is expected to be reached within the next fifty years (5-7). Peak oil production describes the point in time when the maximum rate of petroleum extraction is reached, after that point, there will be a steady decrease in petroleum extraction. Therefore, it is imperative to identify alternative resources for production of chemicals and energy which were previously petroleum based. These are both incredible challenges, but this thesis focuses on the energy crisis.

1.1.2 Alternative and Renewable Energies

As petroleum resources are dwindling, alternative energies must be developed. It is imperative that these alternative energies are derived from a renewable source; otherwise a second energy crisis could occur. There are many avenues for renewable energies originating from natural

resources which can be replenished, such as sunlight, wind, rain, tides, biomass, and geothermal heat. Currently about 16% of global energy consumption comes from renewables: 10% is produced from traditional biomass (mainly for heating), 3.4% from hydroelectricity, and the final 3% is derived from new renewable technologies, such as biofuels, wind, and solar (8).

To identify a strategy which will be most successful in identifying a renewable energy, the largest consumption of energy must be considered. A breakdown of energy consumption with the United States yields the largest fraction of energy consumption to be hydroelectric power at 40%, transportation fuels at 30%, industrial energy at 20%, and residential at 10% (Figure 1.2) (9). When considering the largest amount of energy consumption, hydroelectric power is already derived from renewable sources, thus the next most significant energy consumption is transportation fuels. These fuels are used for cars, planes, and trains. Thus, to make the greatest impact on the energy crisis, major focus has been on attainment of renewable transportation fuels.

There has been some success with electric cars (10), however the existing transportation fuel infrastructure relies on the burning of a carbon-based molecule. Thus, the easiest renewable resource for transportation fuels would be biomass, producing a carbon-based molecule for combustion within an engine. To this end, there has been some success with biofuels: in 2011 the equivalent of 10,000 barrels per day oil was produced as biofuel (4).

1.1.3 Biomass

Biomass is a carbon-based as a renewable energy source derived from garbage, plant, and or waste. Significant research for biofuels focuses on plant biomass. Cellulose is the most abundant component of plant biomass, typically in the range of approximately 35 to 50% of plant

dry weight (11). Cellulose is comprised of long polysaccharide chains of β-linked D-glucose molecules (12) (Figure 1.3). The cellulose fibers are often embedded in a matrix of other structural biopolymers, primarily hemicelluloses and lignin, comprising of 20 to 35% and 5 to 30% of plant dry weight, respectively (11). The hemicellulose is a highly branched chain composed of cross-linked hexose sugars such as D-glucose and also the pentose sugars D-xylose and L-arabinose (12). The lignin consists of phenols and chloroform alcohols which are generally considered unusable (Figure 1.4). Degradation of these biopolymers are the dominant structural feature which limits the rate and extent of biomass utilization (11).

1.1.4 Strategies for Biomass Conversion

The conversion of biomass into usable fuels can be completed either through thermochemical routes, including gasification and pyrolysis, or through biochemical routes, such as microorganism fermentation. Studies have shown that both processes are economically comparable for the production of biofuels (13).

1.1.4.1 Thermochemical Routes

One thermochemical process is gasification: a transformation of raw material into a gaseous mixture through a series of chemical reactions in a controlled amount of air and high temperature (14). The gaseous product is an energy rich mixture of H_2, CO, CO_2, CH_4, C_2H_4, and other impurities such as nitrogen, sulfur, alkali compounds, and tars (15). Different gases or air used in the gasification chemical reaction can yield different products. Specific gasification of the biomass can be used to convert the biomass into a hot synthesis gas, „syn-gas" (16). Syn-gas can then be converted to a variety of fuels including hydrogen by a water gas-shift reaction, a

standard reaction where CO reacts with water to form CO_2 and H_2. Methanol is also produced through methanol synthesis: after the water gas-shift reaction occurs, the CO reacts with water to form CO_2 and H_2, and then there is a hydrogenation of CO_2 to methanol. Alkanes are produced from the biomass by Fischer-Tropsch synthesis using Co-, Fe-, or Ru-based catalysts, the overall chemical reaction shown below (14) .

$$CO + 2H_2 \rightarrow (1/n)C_nH_n + H_2O$$

Another thermochemical process, pyrolysis can produce solid, liquid, and gaseous products (17). Pyrolysis is differentiated from gasification because the reactions are completed in the absence of oxygen. There has been significant success in chemical routes to produce transportation fuels from biomass (18). Some of these successes include a pilot plant built by Honeywell for industrial scale biofuels, which is a 50 million gallon facility to produce green gasoline from biomass.

1.1.4.2 Biochemical Routes

Though thermochemical routes have shown to be successful, the processes require expensive catalysts and high temperature and pressures. Thus, a more sustainable route is biochemical, i.e. microbial fermentation. This process uses lower temperatures, atmospheric pressure, and near-neutral pH. Biotechnology had been declared a key technology for a sustainable chemical industry (1), thus it has been a focus as a technology for sustainable biofuels.

1.1.4.2.1 Pretreatment

In order for the microorganisms to ferment the biomass, it must first be broken down into the simple sugars that can be utilized. This pretreatment step has been cited as one of the most

expensive processes for the biochemical conversion to biofuels (19). The lignin must first be removed from the biomass mixture and the remaining components are treated through a variety of different methods. The cellulose and hemicellulose can be broken down into individual sugars through either steam explosion, mechanical milling, high temperature treatment, and acid/base treatment to release individual sugar molecules (20). These methods are energy intensive and use harsh solvents, thus if researchers are trying to make this process sustainable and economically competitive with petroleum products these are not good methods. Thus, the preferred treatment is enzymatic hydrolysis. An *in vitro* enzyme cocktail is used to digest the hemicellulose and cellulose molecules. The cellulose molecule is broken down by cellulases, such as endoglucanases and cellobiohydrolases. These enzymes cleave the long polysaccharide chain into disaccharide cellobiose molecules. Subsequently, β-glucosidases hydrolyzes the β 1→4 bond of the cellobiose, releasing two D-glucose molecules (21). The hemicellulose fraction consisting of the linked pentose sugars are cleaved by hemicelluloses into singular molecules of D-xylose and L-arabinose (22). The final hydrolysate contains a mixture of sugars: D-glucose, D-xylose, and L-arabinose.

1.1.4.2.2 Biochemical Conversion

Biological processes are attractive because of the variety of specialty chemicals they can produce using a low-energy, low temperature, and low pressure system with non-toxic by-products. The process does not require the use of heavy metals, organic solvents, and strong acids/bases often associated with traditional chemical synthesis, thus making the process sustainably environmentally-friendly (23). Additionally, for many processes, and in the case of bioethanol

production, the system has been finely tuned by natural evolution for efficiency, thus biochemical routes for biofuel production hold significant promise.

Different organisms have been used for biofuel production, including *Saccharomyces cerevisiae, Escherichia coli, Zymomonas mobilis*, and strains of *Clostridium.* Both *S.cerevisiae* and *E.coli* are genetically tractable and desired for most work in microbial catalysis. It is generally agreed that *S.cerevisiae* is the preferred chassis for biofuel production. This microorganism is very industrially robust: tolerant to low pH, resistant to phages, and has a fast doubling time. Additionally, industrial-scale fermentation of this microorganism has been achieved through the beer/wine industry. *S.*cerevisiae is most often used in current industrial-scale bioethanol production and is a goal for many future industrial processes. However, most of the advanced biofuel studies are conducted in *E.coli* because this organism is easier to work with. Future studies of advanced biofuel production will take part in yeast.

1.2 Engineering Microorganisms

The broad field of metabolic engineering has been established to engineer metabolic pathways within a microorganism to produce specialty chemicals. As the field has evolved, a sub-field of pathway engineering has formed. Pathway engineering focuses on optimizing the flux through the pathway itself, wherein metabolic engineering can also involve engineering of the host strain. The projects of this thesis focus on pathway engineering, as the studies have not yet included metabolic engineering strategies for host engineering. Therefore, a detailed analysis of pathway engineering techniques is described, to give a complete background to the pathway engineering methods which were utilized to complete this thesis.

1.2.1 Design and Construction of Pathways

1.2.1.1 Computational Programs for Design

In pathway engineering, there are many challenges, the first being the design of the pathway. Identifying a pathway may require sorting through thousands of possible enzymes and reactions wherein all the parameters have different substrate preferences and kinetic features. Computational algorithms have proven vital in pathway design and have the ability to discover *de novo* pathways by combining enzymes from various sources to produce non-natural products. Several databases containing enzymes and enzymatic reactions have been developed: Kyoto Encyclopedia of Genes and Genomes (KEGG) (24), BRaunschwig ENzyme DAtabase (BRENDA) (25), MetaCyc (26), the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD) (27,28), Retro-Biosynthesis Tool (ReBiT) (29), and the Universal Protein Resource (Uniprot) (30). With these databases, it is possible to identify a staggering number of different pathways for synthesis of the same target compound. Thus a major challenge in computational design is to rank candidate pathways to predict which pathway will have the highest yield of the desired compound.

1.2.1.2 Homologous DNA Recombination Techniques

There are different tools available to construct these large pathways. One set of tools relies on homologous DNA recombination. Of these tools, the DNA assembler method is efficient because it is a one-step *in vivo* pathway construction method. (31). This tool has been directly applied to construct pathways up to 45.3 kb in length with high efficiency. The isothermal Gibson Assembly method developed uses an *in vitro* recombination system to assemble overlapping DNA fragments (32,33). An enzyme cocktail containing T5 exonuclease, Phusion

polymerase, and *Taq* DNA ligase executes overhang chew-back, gap-filling, and ligation all in one-step. Gibson and coworkers reported the construction of a 16.3 kb mouse mitochondrial genome from 600 overlapping 60-mers *in vitro*, demonstrating that this method is a successful approach for multiple-fragment, multiple-size assembly.

1.2.1.3 Restriction Digestion and Ligation Techniques

The BioBrick and BglBrick methods are used to construct a metabolic pathway in a step-by-step manner. Restriction digestions and ligations are implemented with compatible sticky ends generated by a set of specific restriction enzymes. These two approaches allow operon construction, serving as building blocks for pathway assembly. However, the slow manipulation and the existence of a scar in key locations prevent its wide application in pathway construction.

The Golden Gate assembly method has attracted a lot of attention, due to its one-pot, one-step technology which can assemble multiple DNA fragments in 5 minutes. Type IIS enzymes recognize asymmetric sequences and cleave outside of these recognition sites, resulting in overhangs enabling the assembly by ligation. Thus DNA can be inserted into a cloning vector with Type IIS restriction sites; after ligation, the target construct can be established.

1.2.2 Optimization of Pathways

However, pathway engineering is more than just recruiting various enzymes and stringing them together (34). In early heterologous gene expression studies, it was discovered that there was significant growth inhibition in cells over-expressing superfluous genes. This inhibition was attributed to the competition of protein synthesis machinery for essential proteins required for cell growth versus the production of the over-expressed proteins, and the demand for nucleotides

in the replication of the DNA itself. This became known as the metabolic burden and reduction of the metabolic burden has become essential in the pursuit to identify an optimal metabolic pathway (35-38). More than minimizing the metabolic burden, often one or two enzymes in the pathway will not express well or have low activity. This causes a bottleneck or a build-up of intermediates in the pathway, thereby reducing the overall titer. The following are techniques which have been developed to optimize a pathway through optimal protein expression and through altered protein activity.

1.2.2.1 Chromosomal Integration

There are disadvantages to the traditional plasmid-based expression system. Plasmid instability has shown to be a major problem. Due to the metabolic burden on the plasmid-bearing organism, it has a selective disadvantage compared to a plasmid-free cell. Without selection pressure, an entire culture could be dominated by the plasmid-free cell. Integration into the chromosome leads to long-term stability and circumvents the loss of productivity by allele segregation due to plasmid-use. Tyo and coworkers (39) developed a technique called chemical inducible chromosomal evolution (CiChE) to integrate the heterologous genes into the chromosome at multiple sites, with up to 40 consecutive copies.

1.2.2.2 Codon Optimization

The degeneracy of the genetic code is exploited systematically in different ways by different organisms (40,41). Studies have shown that organisms have a bias toward preferred codons and if a gene from an organism with a certain codon bias is heterologously expressed in a host with a different codon bias, the translation of the recombinant protein will be suboptimal. This can be

overcome by codon optimization, wherein the codon sequence of the target gene is altered to match that of the codon bias of the host. Codon optimization is commercially available from DNA synthesis companies and has proven to increase protein expression in multiple studies (42-45).

1.2.2.3 Operons and Tuning Intergenic Regions

Grouping the genes of interest for the pathway into a single polycistronic mRNA can be beneficial to optimize the enzyme expression. This one-step gene expression allows for simplified induction processes and improved regulatory mechanisms. However, the design of the basic operon can greatly affect expression.

Smolke *et al*. (46) showed that altering the mRNA stability by engineering the intergenic regions would increase protein production and thus overall titer. By introducing RNA hairpin turns into the 3‟ and 5‟ intergenic region of the gene of interest, the mRNA was less susceptible to mRNA degradation. This idea of independently controlling gene expression within an operon was further explored by Pfleger *et al*. (47). By creating a library of tunable intergenic regions (TIGRs), the expression of several genes within an operon was simultaneously tuned for optimal expression. The TIGRs contained control elements that include mRNA secondary structures, RNase cleavage sites, and ribosomal binding site (RBS) sequestering sequences. The library of mutated intergenic regions was constructed and applied to the mevalonate pathway. The final mutant product concentration was increased sevenfold compared to the original wild-type pathway.

1.2.2.4 Promoter Strength

Promoters can be mutagenized to achieve precise strength and regulation. These promoters of varying strengths have been applied to exhibit a broad range of genetic control, which can be selected to construct an optimal metabolic pathway. Alper *et al.* (48) used a library of mutant promoters with varying strengths to regulate the expression of the *dxs* gene, which led to the improvement of the volumetric productivity of lycopene. It was found that optimal gene expression did not involve the strongest promoter. A more broadly applicable method was recently established by Du *et al* (49), which relies on the DNA assembler method to automatically combine multiple libraries of mutant promoters of different strengths in a multi-gene pathway. Through a rigorous screening and selection process, the optimal combinations of mutant promoters in the pathway were identified, and involved a balanced expression level. This technique is especially powerful because it involves multiple promoters in a multi-gene pathway, optimizing the expression of all genes simultaneously

1.2.2.5 Protein Engineering

Protein engineering can be used to modify pathways. The previous strategies rely on balancing protein expression, but these strategies cannot overcome innate inefficiencies in enzymes themselves. Thus, engineering the proteins in the pathway can sometimes be the only way to improve the flux through the pathway

As one of the most powerful tools in protein engineering, directed evolution mimics the Darwinian evolutionary process in a test tube and involves iterative rounds of creating genetic diversity followed by selection or screening (50-52) (Fig. 1A). The most common genetic diversity creation methods include error-prone PCR, DNA shuffling, chemical mutagenesis, use

of a mutator strain, and saturation mutagenesis. To identify improved mutants from this genetic diversity, a myriad of screening/selection methods have been developed such as colorimetric assays, colony size-based growth assays, and fluorescence activated cell sorting (FACS). A major advantage of directed evolution is that no prior knowledge of the enzyme structure or mechanism is required to improve enzymatic properties. Another advantage is the ability to mutate the entire enzyme, thus identifying residues distant to the active site that can affect activity through allosteric interactions. However, a major disadvantage of directed evolution is the impossibility of exploring all sequence diversity, even with the most powerful screening or selection method. Additionally, it can be difficult and time consuming to develop a high throughput screening/selection method for a target enzyme property (Figure 1.5a).

Rational design is a knowledge-driven process which uses *a priori* information about enzymes such as sequence or structure. This knowledge is used to make specific, targeted amino acid mutations which are predicted to affect enzymatic properties vital for the efficiency of the desired reaction (Figure 1.5b). This strategy can be valued more than directed evolution because it limits the onerous task of screening the large libraries from directed evolution. In a sequence-based knowledge approach, researchers pursue systematic comparisons of homologous protein sequences to identify possible residues that could alter protein activity. When the three-dimensional crystal structure of the target enzyme or a homologous enzyme is available, a more direct structure-function relationship study between residues can be investigated. Critical residues that are hypothesized to affect the desired function are targeted for mutagenesis. By visualizing the active site, rational design can be used to mutate the large residues to smaller, hydrophobic residues, thus enlarging the active site and allowing a larger substrate to bind.

Various computational tools have been developed to compare the homologous sequences and structural databases to create a mutability map for a target protein (53-55) (Figure 1.5b).

Distinctions between rational design and directed evolution are becoming less clear, as researchers commonly combine these techniques (Figure 1.5c). Though strategies to conjoin the methods vary, one strategy involves a two-step process of directed evolution and then identifying hotspots for rational design, or vice versa if used in *de novo* protein design. Another strategy to combine these techniques is semi-rational design. The power of semi-rational design lies in the reduction of the library size to be screened and an augmentation of the success rate for identifying positive hits (56). Semi-rational design can target specific residues for saturation mutagenesis or target and mutagenize a specific domain that is suspected to have critical function. These intelligent libraries rely on the ability to identify key beneficial mutations through critical structure-function relationships and knowledge of mutational effects on protein folding and activity. This design process harnesses the power of the directed evolution but reduces the onerous screening. Techniques to develop these smart libraries have been diverse and efficient (51,52).

1.2.2.5.1 Engineering Enzyme Specificity or Selectivity

Both rational design and directed evolution have successfully been used to engineer enzyme specificity or selectivity to improve a pathway performance. Often, the promiscuity of enzymes tends to correlate with low catalytic activity and undesired side-reactions. This yields an inefficient pathway with build-up of intermediates and by-products, which can be toxic to the cell. By redesigning proteins for increased specificity or selectivity toward the desired substrate,

less substrate is required for a high titer and fewer by-products are formed. This strategy has proven to be one of the most successful strategies to optimize a pathway through protein design.

One particular example of reducing side-reactions of enzymes is the production of specialty chemical triacetic acid lactone (TAL) (57). The fatty acid synthase (FAS) is a bifunctional enzyme, but was rationally designed through sequence homology and structure-function relationships to inactivate the keto reductase domain. This eliminated the enzyme‟s ability to utilize NADPH, making one function inactive. As a result, the carbon flux was shifted exclusively to the desired product TAL. Another example did not completely eliminate secondary activity, but increased selectivity towards the desired reaction. Nair and Zhao (58) engineered a xylose reductase with preferred selectivity for xylose over arabinose. Screening the error-prone PCR library revealed mutations clustering around the $(\beta/\alpha)_8$-barrel, which led to speculation that this region was involved in substrate recognition. Targeting this secondary structure, the site-saturation mutagenesis led to a final mutant with an increased selectivity from 2.4- to 16.5- fold preference for D-xylose.

Zhang and coworkers (59) engineered a promiscuous KivD for preferred substrate selectivity, which expanded *E.coli* metabolism to produce unnatural C5 to C8 alcohols, used for biofuels. The KivD specificity constant k_{cat}/K_m was engineered through rational design to 40-fold higher preference for 2-keto-4-methylhexanoate than its cognate substrate, 2-ketoisovalerate. This was a 10-fold increase over the wild-type selectivity. The α-keto acid binding pocket was enlarged by mutating proximal residues to smaller, hydrophobic amino acids, allowing binding of the larger substrate. In this same report, LeuA was rationally designed for increased activity for the substrate of interest, 2-keto-3-methylhexanoate. The crystal structure suggested that steric hindrance of the methyl group in the 2-keto-3-methylhexanoate could be relieved by mutating

proximal residues. The mutation, S139G, was identified to increase the k_{cat} 7-fold towards the substrate of interest. The combination of the mutated KivD and LeuA produced several non-natural alcohols, which were not produced in the wild-type pathway.

Just as significant as the substrate, the cofactor $NAD(P)^+$ and NAD(P)H usage can be engineered to increase the overall efficiency of the pathway. Properly designed proteins can reduce the competition for the cofactors between cell metabolism and enzymes in the pathway. An internal cofactor regeneration strategy was incorporated into the xylose utilization pathway to reduce the imbalance between the xylose reductase (XR) and xylitol dehydrogenase (XDH) for biofuel production. In one study, the XDH enzyme was engineered for switched cofactor specificity to $NADP^+$, which would complement the XR cofactor preference of NADPH (60,61), creating an internal cofactor regeneration mechanism. In another study, the XR was altered to have a preference for NADH, to complement the XDH of NAD^+, which led to a 40-fold increase in ethanol productivity over the wild-type enzyme (62).

1.2.2.5.2 Improving Protein Activity

Though improving the substrate/cofactor specificity or selectivity of enzymes can increase flux, there are many other ways to engineer enzymes to increase the overall performance of a pathway. More active enzymes can increase the performance of pathways, and the higher enzymatic activity can be achieved by improving the catalytic efficiency, protein solubility, and stability. Additionally, the pathway can be optimized by increasing the activity of the transport system transferring the extra-cellular substrates into the cell.

Leonard and coworkers (63) engineered two enzymes from the diterpenoid biosynthetic pathway, geranylgeranyl disphosphate synthase (GGPPS) and levopimaradiene synthase (LPS) for

levopimaradiene production. Saturation mutagenesis was applied to possible critical residues of the LPS while error-prone PCR was utilized to engineer the GGPPS. The most improved GGPPS (S239C/G295D) and LPS (M593I/Y700F) were cloned into the full pathway and when combined with metabolic engineering strategies, resulted in a 2,600-fold increase in levopimaradiene production. The mutations in the GGPPS were hypothesized to affect the GGPPS catalysis by improving the binding efficiency of the magnesium ions needed for substrate anchoring.

Increasing the overall enzyme activity through traditional directed evolution was eloquently illustrated by Atsumi and Liao in a study focusing on 1-propanol and 1-butanol synthesis in *E. coli* (64). A combination of error-prone PCR and DNA shuffling through six rounds of directed evolution identified a top candidate, called CimaA3.7. The k_{cat}/K_m of CimaA3.7 for acetyl-CoA was 6.7 times greater than the wild-type enzyme at 30 ºC. The enzyme activity was screened at varying temperatures and CimA3.7 showed higher activity than wild-type at all temperatures tested. Once re-inserted into the pathway, this enzyme yielded a 9.2-fold improvement in 1-propanol production and a 21.9-fold improvement in 1-butanol production.

Transporters can be the greatest limiting factor in a pathway. However, engineering these proteins has been an underutilized strategy for improving the performance of pathways. Lack of three-dimensional crystal structure has limited the engineering of these proteins to error-prone and directed evolution. Young and coworkers (65) engineered xylose transporters with an improved growth rate of up to 70% on xylose as the sole carbon source. In this engineering effort, the transporters with increased affinity towards xylose were also discovered to have a lower V_{max}.

1.2.3 Engineering Biochemical Pathways for Biofuels

As previously mentioned, hexose (D-glucose) is a main constituent of biomass; however pentose sugars such as D-xylose and L-arabinose from the hemicellulose fraction can contribute to a large component of the biomass. *S.cerevisiae* and *E.coli* cannot utilize these pentose sugars. Thus, a significant amount of engineering has transpired to design these microorganisms to efficiently utilize these sugars to produce bio-ethanol. Bio-ethanol has emerged as a first-generation biofuel. However, it is well established that this biofuel is inefficient, with low heating properties and inefficiencies. Thus, production of advanced biofuels has now become a focus: butanol, biodiesel, alkanes, and hydrocarbons. Though many studies recently focused on producing advanced biofuels, either through *E.coli* or *S.cerevisiae*, no study has linked pentose utilization to advanced biofuel production.

1.2.3.1 Pentose Sugar Utilization and Bioethanol Production

Optimizing the pentose sugar utilization pathway has two main thrusts: engineering transporters to transport the sugars into the cell without incurring a diauxic shift and optimizing the enzymes which convert the sugars to ethanol.

1.2.3.1.1 Engineering Transporters

The diauxic shift describes a phenomenon wherein if a mixture of sugars is present, such as D-glucose, D-xylose, and L-arabinose, the microorganisms will preferentially utilize the glucose molecules, subsequently the xylose and finally the arabinose. This sequential sugar utilization greatly decreases overall efficiency of the process (12,66-68). Many metabolic engineering techniques have been used to overcome this phenomenon. Specifically, pathway engineering

strategies focus on engineering the transporters themselves. It has been shown that using pentose-specific transporters can reduce the diauxic shift, because the sugars will not utilize the same transporters as the glucose molecules, allowing for simultaneous transport. These projects identified pentose-specific transporters and/or engineered the transporters for higher activity (65,69).

Another strategy for reducing the diauxic shift is to utilize cellobiose as a carbon source. Cellobiose is a glucose dimer and can be transported into the cell by a recently discovered cellodextrin transporter (70). Previous work has reduced the diauxic shift by co-expression the cellodextrin transporter and a β-glucosidase with the xylose utilization genes (71,72).

1.2.3.1.2 Xylose and Arabinose Utilization

One of the most successful strategies for xylose utilization is the heterologous expression of the fungal xylose utilizing pathway. The xylose is reduced to xylitol by a xylose reductase (XR). Then the xylitol is oxidized to D-xylulose by a xylitol dehydrogenase (XDH). Finally, a xylulose kinase (XKS) converts the D-xylulose to D-xylulose-5-P, which is then shuttled to the pentose phosphate pathway for ethanol production (Figure 1.6). The arabinose pathway is similar: the arabinose is reduced to L-arabitol by the XR, an L-arabitol 4-dehydrogenase (LAD) reduces the arabitol to L-xylulose, which is then converted to D-xylulose by a L-xylulose reductase (LXR). Though the xylose pathway has shown to be efficient (49), this arabinose pathway is not.

The imbalance of the cofactor usage between the enzymes is suspected to be the greatest limiting factor (73). Strategies to over this limitation were previously discussed in the chapter, involving engineering enzymes for altered cofactor specificity. Thus internal cofactor regeneration mechanisms were established. Significant work has looked at balancing the promoters of the

xylose utilization enzymes. These studies have shown that a high XR expression is important for an efficient pathway (49).

1.2.3.2 Advanced Biofuels

Though bioethanol has been marginally successful as a biofuels, more research is focusing on advanced biofuels, such as butanol, biodiesel, alkanes/alkenes, and hydrocarbons. These advanced biofuels have more similar properties to petroleum-based fuels than ethanol, and are termed „drop in" fuels because they can readily be used by current engines in cars and planes.

Recent work has focused on the Ehrlich pathway or the 2-keto-acid pathway for production of butanol and isobutanol. Intermediates of the amino acid pathway, keto acids, are decarboxylated into aldehydes and then reduced to the products. Isobutanol is perhaps one of the products produced by microbes which are closest to industrial use, with a high energy content and extensive branching, which gives it a higher octane number. The octane number is a measure of a fuel"s resistance to knocking in spark ignition engines (74). The pathway was introduced into *E.coli*, the promiscuous decarboxylase and aldehyde dehydrogenase were engineered for increased substrate specificity in order for the final production of the non-natural carbon chains (59).

Coupling microbial catalysis and chemical catalysis has been shown successful with isoprenoid based biofuels, such as farnesane (75) and bisabolane (76). The intermediates farnesene and bisabolene are produced microbially and then chemically converted to farnesane and bisabolane, which are currently being investigated for as alternative biosynthetic diesel. Plants are known to produce terpenes, thus, plant isoprenoid synthases were screened for efficient conversion of farnesyl diphosphate from the mevalonate pathway into bisabolene (76).

The cell membrane mainly consists of fatty acids. Certain organisms often overproduce fatty acids which are used for energy storage. The long hydrocarbon fatty acyl chain is energy rich, and can be used as a biofuel. Fatty acids are biosynthesized naturally by either a large multienzyme system the fatty acid synthase or small sub-unit proteins. The multienzyme system uses malonyl-CoA as a building block. Fatty acyl chains are elongated on acyl carrier proteins (ACPs) through repeated cycles of decarboxylative condensation, β-keto reduction, dehydration and enol reduction (74).

Thioesterases have been used to release the long-chain fatty acids to produce free fatty acid ethyl ester (FAEEs) (77). Combined with the deletion of genes in the fatty-acid degradation pathway (β-oxidation), free fatty acids have been produced at high titers (1.2 g/L) (78).

Alkane biosynthesis has recently been a major research focus. Derived from fatty acids, alkanes are a predominant component of diesel. Alkane biosynthetic genes were recently discovered in cyanobacteria (79). The acyl-ACP reductase (AAR) was used to reduce the acyl-ACP to aldehydes which were then converted into alkanes by an aldehyde decarboxylase (ADC). A combination of expression of these enzymes in *E.coli* produced alkanes from C_{13}-C_{17} with final yield of 300 mg/L. A figure summarizing some of the advanced biofuel pathways is available as Figure 1.7. A more detailed discussion of advanced biofuels is available in Chapter 3.

1.3 Overview of Projects

The following projects in this thesis focus on the design, construction, and optimization of biochemical pathways for production of biofuels. The second chapter entitled "Optimization of a Cellobiose Pathway through Directed Evolution" optimizes an already established cellobiose utilization pathway. Though many of the techniques previously discussed to optimize the

pathway could have been used, in this project we developed a new strategy to engineer two enzymes simultaneously through directed evolution to increase the specific growth rate on cellobiose. This is one of the few examples of directed evolution on the pathway scale. Two rounds of genetic diversity identified a mutant pathway which increases the specific growth rate on cellobiose by 41% and increase ethanol productivity by 74%. The enzymes were individually characterized to identify how the mutations affected the activity. The β-glucosidase enzyme was engineered for increased substrate specificity and activity. Simultaneously, the cellodextrin transporter was engineered for higher activity. This project illustrates for the first time that multiple enzymes have been engineered in a single pathway simultaneously in *S.cerevisiae*.

The third chapter involves the design and construction of a novel biochemical route for the production of free fatty acid ethyl esters (FAEEs). The fatty acid synthase (FAS) from *Brevibacterium ammoniagenes* is used to heterologously produce excess fatty acid C_{16}. This enzyme is then coupled with a wax ester/ diacylglycerol acyltransferase (WS/DGAT) to produce the FAEEs. Plasmid construction has been completed. Initial single enzyme assays were completed for both the FAS and the WS/DGAT in *S.cerevisiae*. The WS/DGAT was shown to be active, however, detection of C_{16} produced by the FAS was not detected. Future work in this project will involve detection of FAEE and metabolic engineering to improve FAEE production.

1.4 Conclusions

Energy demand is increasing at an alarming rate, with an increase of 2.5% in 2011. As peak oil extraction will be reached in the next 50 years, alternative energies must be identified. One of the most efficient alternative energies is biofuels. Engineering microorganisms to produce these biofuels is promising. The techniques to optimize a metabolic pathway which produces these

biofuel-type molecules are wide and diverse. Specifically, pathway engineering offers many different options to be able to optimize the flux through the pathway. Though engineering the protein expression has proven successful, often there are limitations and inefficiencies of the enzymes which need to be engineered. Thus, significant protein engineering tools and methods have developed. This thesis focuses on developing a new method for the simultaneous engineering of multiple enzymes, which was applied to a cellobiose utilization pathway. These same techniques can be applied to novel pathways for the production of advanced biofuels, but first those pathways must be designed and constructed. In this thesis, a biodiesel pathway is designed and constructed. Future work will optimize the constructed pathway.

1.5 References

1. Gavrilescu, M. and Chisti, Y. (2005) Biotechnology-a sustainable alternative for chemical industry. *Biotechnol. Adv.*, **23**, 471-499.
2. Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Jr., Hallett, J.P., Leak, D.J., Liotta, C.L. *et al.* (2006) The path forward for biofuels and biomaterials. *Science*, **311**, 484-489.
3. US Energy Information Administration. (2011) World Energy Consumption.
4. British Petroleum. (2012) BP Statistical Review of World Energy June 2012.
5. Hughes, L. and Rudolph, J. (2011) Future world oil production: growth, plateau, or peak? *Curr. Opin. Environ. Sustainability*, **3**, 225-234.
6. Kjarstad, J. and Johnsson, F. (2009) Resources and future supply of oil. *Energy Policy*, **37**, 441-464.
7. Sorrell, S., Speirs, J., Bentley, R., Brandt, A. and Miller, R. (2010) Global oil depletion: A review of the evidence. *Energy Policy*, **38**, 5290-5295.
8. Renewable Energy Policy Network for the 21st Century. (2011) Renewables 2011: Global Status Report.
9. U.S. Energy Information Administration. (2012) Annual Energy Review 2011.
10. Tran, M., Banister, D., Bishop, J.D.K. and McCulloch, M.D. (2012) Realizing the electric-vehicle revolution. *Nat. Clim. Change*, **2**, 328-333.
11. Lynd, L.R., Weimer, P.J., van Zyl, W.H. and Pretorius, I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, **66**, 506-577.
12. Aristidou, A. and Penttilä, M. (2000) Metabolic engineering applications to renewable resource utilization. *Curr. Opin. Biotechnol.*, **11**, 187-198.
13. Wright, M.M. and Brown, R.C. (2007) Comparative economics of biorefineries based on the biochemical and thermochemical platforms. *Biofuels, Bioprod. Biorefin.*, **1**, 49-56.
14. Huber, G.W., Iborra, S. and Corma, A. (2006) Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. *Chem. Rev.*, **106**, 4044-4098.
15. Damartzis, T. and Zabaniotou, A. (2011) Thermochemical conversion of biomass to second generation biofuels through integrated process design-A review. *Renewable Sustainable Energy Rev.*, **15**, 366-378.
16. NREL. (2006) NREL Leads the Way: From Biomass to Biofuels.
17. Mohan, D., Pittman, C.U. and Steele, P.H. (2006) Pyrolysis of wood/biomass for bio-oil: A critical review. *Energy & Fuels*, **20**, 848-889.
18. Demirbas, M.F. and Balat, M. (2006) Recent advances on the production and utilization trends of bio-fuels: A global perspective. *Energy Convers. Manage.*, **47**, 2371-2381.
19. Harmsen, P.H., Wouter. Bermudez, Laura. Bakker, Robert. (2010) BioSynergy Report 1184.
20. Mielenz, J.R. (2001) Ethanol production from biomass: technology and commercialization status. *Curr. Opin. Microbiol.*, **4**, 324-329.
21. Galazka, J.M., Tian, C.G., Beeson, W.T., Martinez, B., Glass, N.L. and Cate, J.H.D. (2010) Cellodextrin transport in yeast for improved biofuel production. *Science*, **330**, 84-86.
22. Maris, A.J.A.v., Abbott, D.A., Bellissimi, E., Brink, J.v.d., Kuyper, M., Luttik, M.A.H., Wisselink, H.W., Scheffers, W.A., Dijken, J.P.v. and Pronk, J.T. (2006) Alcoholic

fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Antonie van Leeuwenhoek*, **90**, 391-418.
23. Du, J., Shao, Z.Y., Zhao, H.M. (2011) Engineering microbial factories for synthesis of value-added products. *J. Ind. Microbiol. Biotechnol.*, **38**, 873-890.
24. Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27-30.
25. Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. and Schomburg, D. (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.*, **32**, D431-D433.
26. Caspi, R., Foerster, H., Fulcher, C.A., Hopkinson, R., Ingraham, J., Kaipa, P., Krummenacker, M., Paley, S., Pick, J., Rhee, S.Y. *et al.* (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.*, **34**, D511-D516.
27. Hou, B.K., Ellis, L.B.M. and Wackett, L.P. (2004) Encoding microbial metabolic logic: predicting biodegradation. *J. Ind. Microbiol. Biotechnol.*, **31**, 261-272.
28. Ellis, L.B.M., Roe, D. and Wackett, L.P. (2006) The University of Minnesota Biocatalysis/Biodegradation Database: the first decade. *Nucleic Acids Res.*, **34**, 517-521.
29. Prather, K.L.J. and Martin, C.H. (2008) *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.*, **19**, 468-474.
30. Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H.Z., Lopez, R. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, 187-191.
31. Shao, Z.Y., Zhao, H. and Zhao, H.M. (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.*, **37**.
32. Ramon, A. and Smith, H.O. (2011) Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett.*, **33**, 549-555.
33. Gibson, D., Smith, H., Hutchison, C., Venter, J.C. and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods*, **7**, 901-U905.
34. Khosla, C. and Keasling, J.D. (2003) Timeline - Metabolic engineering for drug discovery and development. *Nat. Rev. Drug Discovery*, **2**, 1019-1025.
35. Bentley, W.E., Mirjalili, N., Andersen, D.C., Davis, R.H. and Kompala, D.S. (1990) Plasmid-encoded protein - The principle factor in the metabolic burden associated with recombinant bacteria. *Biotechnol. Bioeng.*, **35**, 668-681.
36. Vind, J., Sorensen, M.A., Rasmussen, M.D. and Pedersen, S. (1993) Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes - expression from reporter genes does not always reflect functional messenger-RNA levels. *J. Mol. Biol.*, **231**, 678-688.
37. Birnbaum, S. and Bailey, J.E. (1991) Plasmid presence changes the relative levels of many host-cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnol. Bioeng.*, **37**, 736-745.
38. Glick, B.R. (1995) Metabolic load and heterologous gene-expression. *Biotechnol. Adv.*, **13**, 247-261.
39. Tyo, K.E.J., Ajikumar, P.K. and Stephanopoulos, G. (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat. Biotechnol.*, **27**, 760-115.

40. Gustafsson, C., Govindarajan, S. and Minshull, J. (2004) Codon bias and heterologous protein expression. *Trends Biotechnol.*, **22**, 346-353.
41. Ernst, J.F. (1988) Codon usage and gene expression. *Trends Biotechnol.*, **6**, 196-199.
42. Kink, J.A., Maley, M.E., Ling, K.Y., Kanabrocki, J.A. and Kung, C. (1991) Efficient expression of the paramecium calmodulin gene in *Escherichia coli* after 4 TAA-to-CAA changes through a series of polymerase chain-reactions. *J. Protozool.*, **38**, 441-447.
43. Nambiar, K.P., Stackhouse, J., Stauffer, D.M., Kennedy, W.P., Eldredge, J.K. and Benner, S.A. (1984) Total synthesis and cloning of a gene coding for the ribonuclease-S protein. *Science*, **223**, 1299-1301.
44. Tokuoka, M., Tanaka, M., Ono, K., Takagi, S., Shintani, T. and Gomi, K. (2008) Codon optimization increases steady-state mRNA levels in *Aspergillus oryzae* heterologous gene expression. *Appl. Environ. Microbiol.*, **74**, 6538-6546.
45. Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D. and Keasling, J.D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.*, **21**, 796-802.
46. Smolke, C.D., Carrier, T.A. and Keasling, J.D. (2000) Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Appl. Environ. Microbiol.*, **66**, 5399-5405.
47. Pfleger, B.F., Pitera, D.J., D Smolke, C. and Keasling, J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.*, **24**, 1027-1032.
48. Alper, H., Fischer, C., Nevoigt, E. and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 12678-12683.
49. Du, J., Yuan, Y., Si, T., Lian, J. and Zhao, H. (2012) Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res.*, **40**, e142.
50. Rubin-Pitel, C.M.-H.C., W. Chen, and H. Zhao. (ed.) (2006) *Directed evolution tools in bioproduct and bioprocess development*. Elsevier Science, New York, NY.
51. Cobb, R.E., Si, T. and Zhao, H. (2012) Directed evolution: an evolving and enabling synthetic biology tool. *Curr. Opin. Chem. Biol.*, **16**, 285-291.
52. Wang, M., Si, T. and Zhao, H. (2012) Biocatalyst development by directed evolution. *Bioresour. Technol.*, **115**, 117-125.
53. Pavelka, A., Chovancova, E. and Damborsky, J. (2009) HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.*, **37**, W376-W383.
54. Pleiss, J. (2011) Protein design in metabolic engineering and synthetic biology. *Curr. Opin. Biotechnol.*, **22**, 611-617.
55. Damborsky, J. and Brezovsky, J. (2009) Computational tools for designing and engineering biocatalysts. *Curr. Opin. Chem. Biol.*, **13**, 26-34.
56. Quin, M.B. and Schmidt-Dannert, C. (2011) Engineering of biocatalysts - from evolution to creation. *ACS Catal.*, **1**, 1017-1021.
57. Zha, W., Shao, Z., Frost, J.W. and Zhao, H. (2004) Rational pathway engineering of type I fatty acid synthase allows the biosynthesis of triacetic acid lactone from D-glucose *in vivo*. *J. Am. Chem. Soc.*, **126**, 4534-4535.
58. Nair, N.U. and Zhao, H. (2008) Evolution in reverse: engineering a D-xylose-specific xylose reductase. *Chembiochem*, **9**, 1213-1215.

59. Zhang, K., Sawaya, M.R., Eisenberg, D.S. and Liao, J.C. (2008) Expanding metabolism for biosynthesis of nonnatural alcohols. *Proc. Natl. Acad. Sci. USA*, **105**, 20653-20658.
60. Krahulec, S., Klimacek, M. and Nidetzky, B. (2009) Engineering of a matched pair of xylose reductase and xylitol dehydrogenase for xylose fermentation by *Saccharomyces cerevisiae*. *Biotechnol. J.*, **4**, 684-694.
61. Matsushika, A., Watanabe, S., Kodaki, T., Makino, K., Inoue, H. and *et al.* (2008) Expression of protein engineered $NADP^+$-dependent xylitol dehydrogenase increases ethanol production from xylose in recombinant *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.*, **81**, 243-255.
62. Runquist, D., Hahn-Hagerdal, B. and Bettiga, M. (2010) Increased ethanol productivity in xylose-utilizing *Saccharomyces cerevisiae* via a randomly mutagenized xylose reductase. *Appl. Environ. Microbiol.*, **76**, 7796-7802.
63. Leonard, E., Ajikumar, P.K., Thayer, K., Xiao, W.H., Mo, J.D. and *et al.* (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci. USA*, **107**, 13654-13659.
64. Atsumi, S. and Liao, J.C. (2008) Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. *Appl. Environ. Microbiol.*, **74**, 7802-7808.
65. Young, E.M., Comer, A.D., Huang, H. and Alper, H.S. (2012) A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. *Metab. Eng.*, **14**, 401-411.
66. Santangelo, G.M. (2006) Glucose signaling in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews*, **70**, 253-+.
67. Rolland, F., Winderickx, J. and Thevelein, J.M. (2002) Glucose-sensing and -signalling mechanisms in yeast. *FEMS Yeast Res.*, **2**, 183-201.
68. Gancedo, J.M. (1998) Yeast carbon catabolite repression. *Microbiology and Molecular Biology Reviews*, **62**, 334-+.
69. Du, J., Li, S. and Zhao, H. (2010) Discovery and characterization of novel D-xylose-specific transporters from *Neurospora crassa* and *Pichia stipitis*. *Mol. Biosyst.*, **6**, 2150-2156.
70. Galazka, J.M., Tian, C., Beeson, W.T., Martinez, B., Glass, N.L. and Cate, J.H.D. (2010) Cellodextrin transport in yeast for improved biofuel production. *Science*, **330**.
71. Ha, S.J., Galazka, J.M., Kim, S.R., Choi, J.H., Yang, X., Seo, J.H., Glass, N.L., Cate, J.H. and Jin, Y.S. (2011) Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 504-509.
72. Li, S., Du, J., Sun, J., Galazka, J.M., Glass, N.L., Cate, J.H.D., Yang, X. and Zhao, H. (2010) Overcoming glucose repression in mixed sugar fermentation by co-expressing a cellobiose transporter and a beta-glucosidase in Saccharomyces cerevisiae. *Mol. Biosyst.*, **6**.
73. Van Vleet, J.H. and Jeffries, T.W. (2009) Yeast metabolic engineering for hemicellulosic ethanol production. *Curr. Opin. Biotechnol.*, **20**, 300-306.
74. Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B. and Keasling, J.D. (2012) Microbial engineering for the production of advanced biofuels. *Nature*, **488**, 320-328.
75. Renninger, N.M., D. (2008). Fuel Compositions Comprising Farnesane and Method Of Making and Using Same. Patent Number: 20080098645

76. Peralta-Yahya, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D. and Lee, T.S. (2011) Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.*, **2**, 483.
77. Lennen, R.M., Braden, D.J., West, R.A., Dumesic, J.A. and Pfleger, B.F. (2010) A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in *Escherichia coli* and catalytic conversion to alkanes. *Biotechnol. Bioeng.*, **106**, 193-202.
78. Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S.B. and Keasling, J.D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, **463**, 559-562.
79. Schirmer, A., Rude, M.A., Li, X., Popova, E. and del Cardayre, S.B. (2010) Microbial biosynthesis of alkanes. *Science*, **329**, 559-562.
80. Rubin, E. (2008) Genomics of cellulosic biofuels. *Nature*, **454**, 841-845.
81. Zhang, F., Rodriguez, S. and Keasling, J.D. (2011) Metabolic engineering of microbial pathways for advanced biofuels production. *Curr. Opin. Biotechnol.*, **22**, 775-783.

1.6 Figures

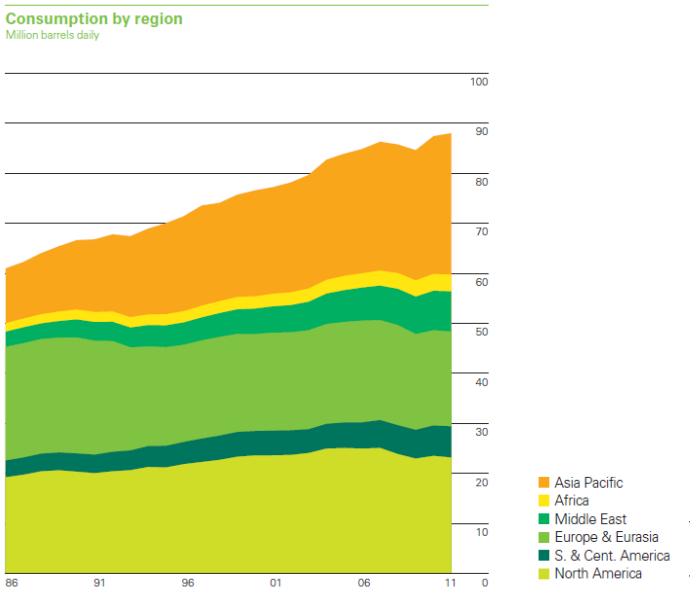


Figure 1.1: Energy consumption by region as predicted by BP in 2012 (4).

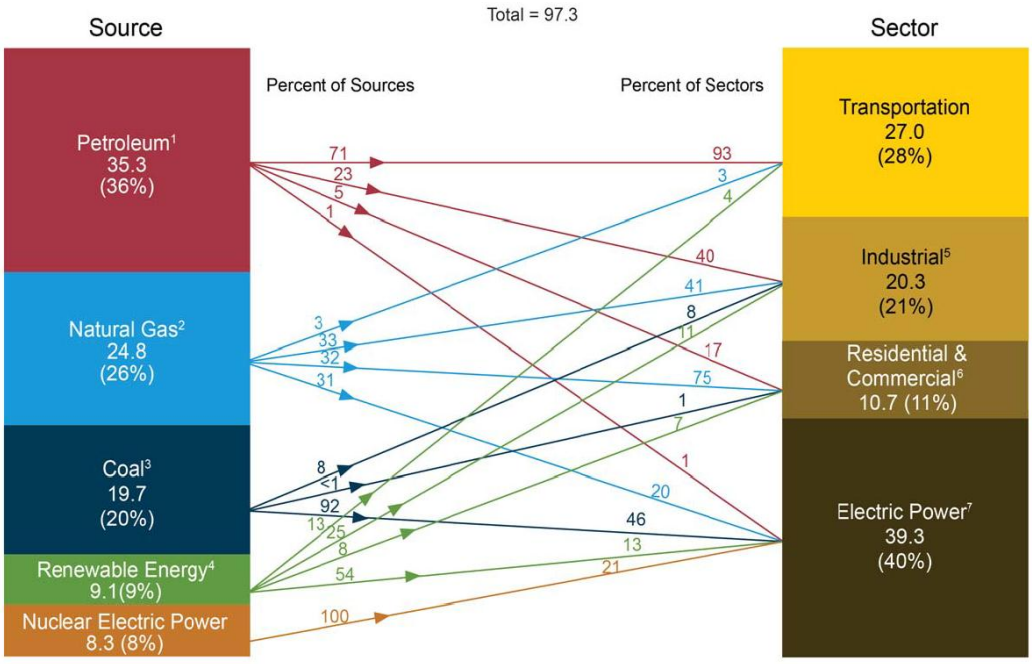


Figure 1.2: Breakdown of energy resources and consumption in the United States in 2011. Each number on the line indicates the percentage of that component (9).

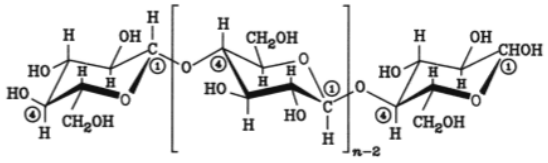


Figure 1.3: A cellulose molecule.

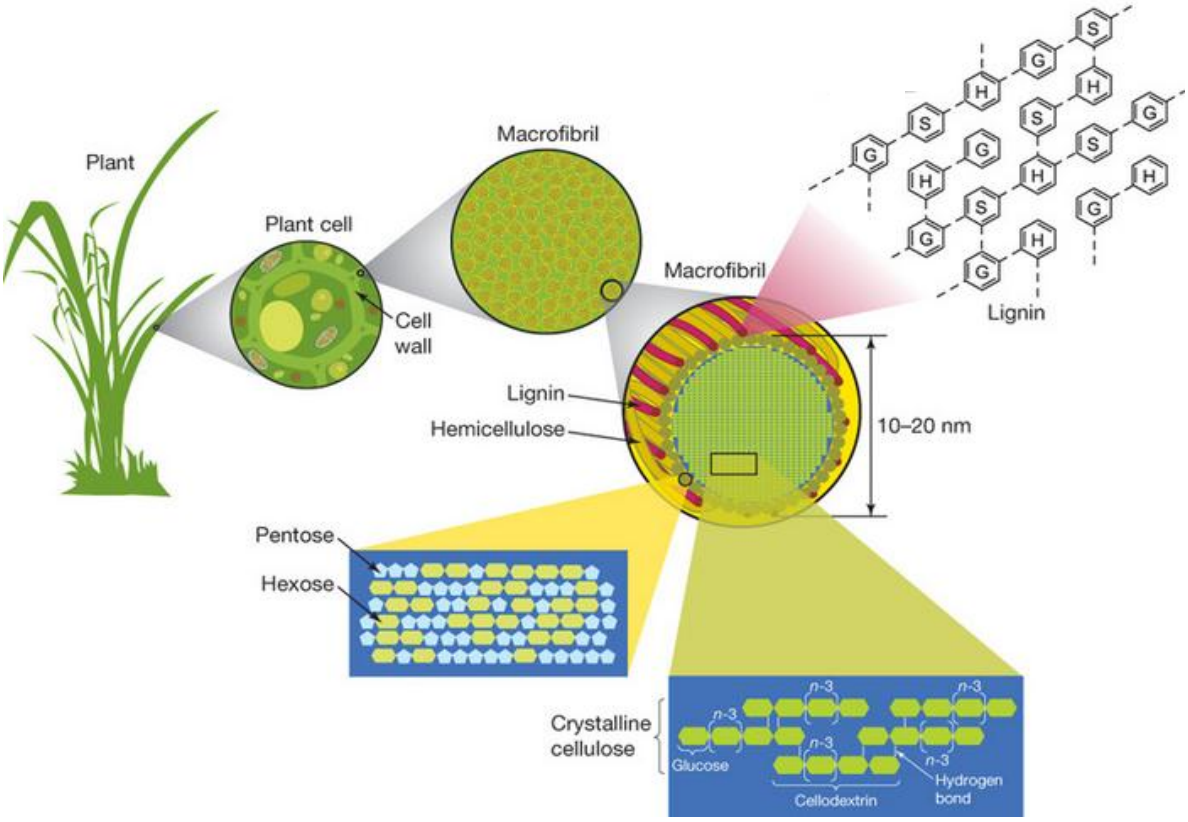


Figure 1.4: Biomass composition (80).

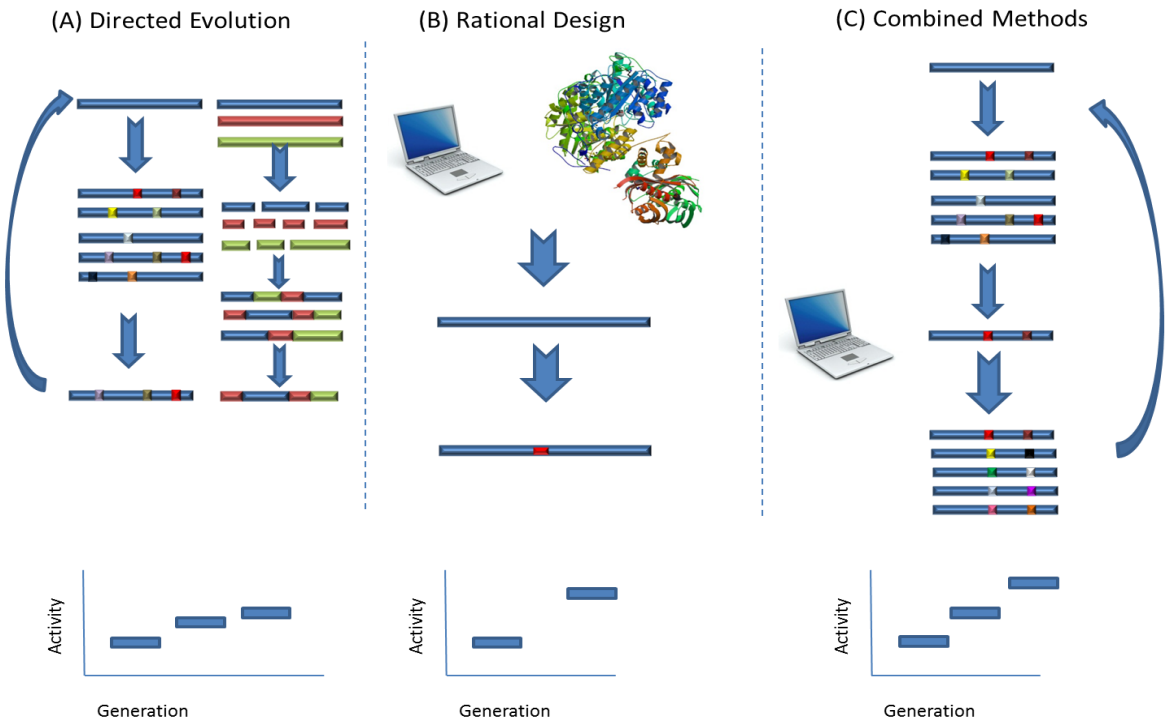


Figure 1.5: Tools used for protein engineering.

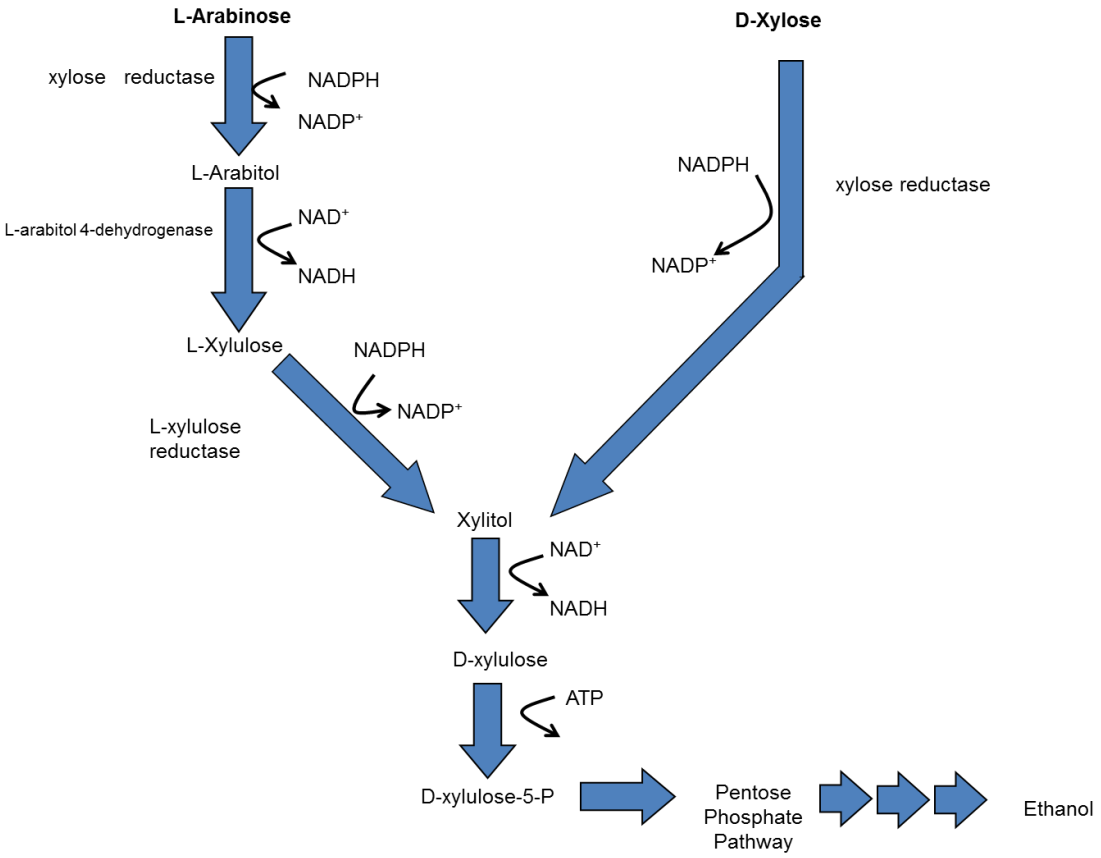


Figure 1.6: Fungal heterologous pentose sugar utilization.

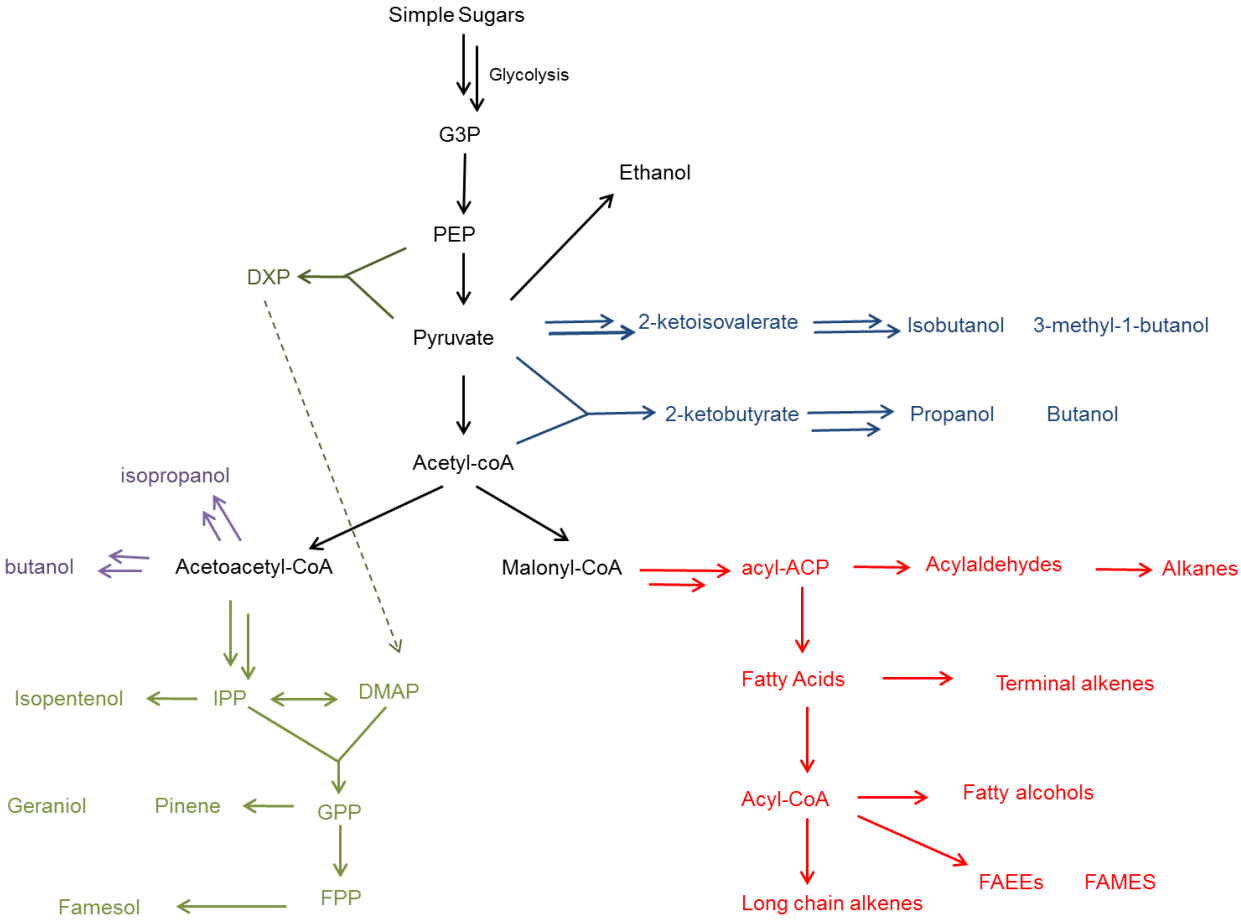


Figure 1.7: Pathways for different advanced biofuels (81).

Chapter 2: Optimization of a Cellobiose Utilization Pathway through Directed Evolution

2.1 Background

The saccharification of lignocellulosic biomass into fermentable sugars is recognized as one of the most expensive operations in biofuel process economics (1,2). Hydrolysis of the polysaccharide constituents of biomass into simple sugars requires an exorbitant amount of enzymes: endoglucanases (EC 3.2.1.4) and exoglucanases (EC3.2.1.74) that collectively break down the cellulose into cellobiose, a β-1,4-glucose disaccharide. Finally, β-glucosidases (EC 3.2.1.21) hydrolyze cellobiose to glucose which is subsequently fermented (3). To reduce the requirement of these *in vitro* hydrolytic enzymes, recent research has investigated the option of utilizing cellobiose itself as a fermentable sugar. Reducing the price of pretreatment will lower the cost of biofuel production, making biofuels more economically competitive with petroleum-derived fuels (1,2).

Three main strategies for heterologous cellobiose utilization have recently developed. The first strategy involves cell-surface display of extracellular β-glucosidases (4-7); the cell metabolizes two glucose molecules resulting from the extracellular hydrolysis of the cellobiose disaccharide. Another approach relies on heterologously expressed transporters and a cellobiose phosphorylase (8-10). The cellobiose is transported into the cell and the phosphorylase cleaves the disaccharide with an inorganic phosphate, producing a glucose molecule and α-glucose-1-phosphate. The final strategy relies on heterologous expression of a cellodextrin transporter and intracellular β-glucosidase (11-13). The intracellular cellobiose is hydrolyzed into two glucose molecules which are subsequently phosphorylated to glucose-6-phosphate (Figure 2.1). Previous studies in our laboratory have investigated optimizing the hydrolytic cellobiose utilization pathway through combinatorial transcriptional engineering (14), however this optimization did not consider

protein activity. Based on the promiscuity of the proteins involved in the hydrolytic pathway, we sought to apply protein engineering methods to improve the pathway.

Current methods for protein engineering often involve isolating the protein from the context of the pathway and engineering it separately. In this study, we have investigated a complementary strategy for protein engineering wherein we simultaneously modified multiple proteins in the context of the pathway (Figure 2.2). Advantages of this strategy allow for simple and efficient library creation to identify pathways which can include mutations in each protein synergistically increasing the desired activity. This approach can be widely applicable to any pathway where a high-throughput screening/selection method for the desired phenotype is available. The implementation of this directed evolution strategy will further enhance researcher‟s abilities to optimize pathways, introducing a next step in metabolic engineering involving pathway-driven approaches for protein engineering.

Here we report the first example of simultaneously engineering multiple proteins, β-glucosidase (*gh1-1)* [GenBank Accession number XM_951090] and cellodextrin transporter (*cdt-1)* [GenBank Accession number XM_958708] in *Saccharomyces cerevisiae* for biofuel production. Through this directed evolution, key mutations were found in both of the proteins which synergistically improved the overall cellobiose utilization by 42%. These mutations were directly linked to improved activity or altered substrate specificity of the proteins. Though improving the cellobiose utilization for biofuel production is of interest, these results also augment recent research in β-glucosidase substrate specificity (15-18) and continued research in transporter engineering (10,19-22).

2.2 Results

2.2.1 Library Construction and Screening

The genes *gh1-1* and *cdt-1* were independently mutagenized through error-prone PCR and co-assembled into the single copy pRS-KanMX plasmid under the PGK and TEF constitutive promoters, respectively. The industrial *S.cerevisiae* strain Still Spirits (Classic) Turbo Distiller‟s Yeast purchased from Homebrew Heaven (Everett, WA). This strain was used to assemble the pathway via the DNA Assembler method (23). The genetic diversity for each gene was 2-3 nucleotide mutations per kilobase in a library of over 10^4 mutants. The growth rate on cellobiose as the sole carbon source was utilized as an assessment of pathway improvement, therefore screening was based on colony size from agar plates (14,24). Large colonies from the library were visibly distinct from the wild-type parent pathway (Figure 2.3) and were selected for quantitative analysis. Plasmids containing the mutant pathways exhibiting increased growth rates on cellobiose in oxygen limited conditions were retransformed into fresh yeast cells to confirm that the increased growth rate was caused by the mutated pathway and not from genome modification (Figure 2.4). The first round of error-prone PCR yielded a mutant pathway (R1) which was used as template for the second round of error-prone PCR. The second round of error-prone PCR resulted in the mutant pathway R2. A third round of mutagenesis was performed but did not result in an improved phenotype.

2.2.2 Characterization of the Mutant Cellobiose Utilizing Pathways

The first mutant pathway, identified as (R1), resulted in the amino acid mutation L173H in the β-glucosidase and a D433G mutation in the cellodextrin transporter. The second round of mutagenesis (R2) resulted in a strain with additional mutations: H23L mutation in the β-

glucosidase and a C82S mutation in the cellodextrin transporter. R1 and R2 exhibited improved cellobiose growth, cellobiose utilization, and ethanol productivity compared to the wild-type cellobiose pathway in oxygen limited conditions (Figure 2.5). There was a 42% increase in specific growth rate in the final mutant (0.075 h^{-1}) compared to the wild-type (0.053 h^{-1}). Metabolite analysis for the engineered pathways presented a 54% increase in cellobiose consumption and a 74% increase in ethanol productivity (Table 2.1). Glucose production was minimal at less than 10 g/L and was found to be transient, being re-consumed quickly (Figure 2.6). The final mutant pathway, R2, consumed 80 g/L cellobiose in about 27 hours with an ethanol productivity of 1.03 g/(L·h) (Figure 2.5 and Table 2.1).

The double and single mutations from each round were isolated and compared for specific growth rate and ethanol productivity (Figure 2.7) to determine which mutations had the most significant effect on the activity of the pathway. Considering the mutation effects of each gene, the L173H mutation in the β-glucosidase is required in the pathway for a high growth rate. In the cellodextrin transporter, the C82S mutation is also needed for higher growth rate and ethanol productivity. It is the combination of these mutations which confers a high growth rate. However, only when all four mutations are present, two from each protein, that the highest growth rate is achieved.

2.2.3 Activity Assays

2.2.3.1 β-glucosidase Assay

The enzymes were assayed *in vitro* to better identify how the mutations affected the activity of the individual enzymes. Initial tests on the β-glucosidase assayed the colorimetric reaction with *p*-nitrophenyl β-D-glucopyranoside (*p*-NPG), a common assay used for β-glucosidase activity

(15,17). Compared to the wild-type, the L173H mutation conferred a reduction in hydrolytic activity towards *p*-NPG by 30%. The combined L173H+H23L increased the activity slightly from the single L173H mutation. Meanwhile, the single H23L mutation had similar activity to the wild-type on *p*-NPG. Next, we assayed the activity towards cellobiose by enzymatically quantifying the glucose produced by a timed reaction of β-glucosidase with cellobiose. Unlike the *p*-NPG assay, the L173H mutation increased the specific activity towards cellobiose by 90% compared to the wild-type. The added H23L mutation conferred a 30% increase in cellobiose activity. The isolated H23L mutation had only slightly higher activity than the wild type enzyme. Overall, the β-glucosidase double mutant had 2.5-fold higher cellobiose hydrolytic activity than the wild-type enzyme, illustrating additive effects with the combined L173H and H23L (Table 2.2).

2.2.3.2 Cellodextrin Assay

The specific activity of the cellodextrin transporter was measured through a modified oil-stop assay based on the rate of radiolabeled sugar uptake (11,25). The mutation D433G from R1 conferred a small increase in cellodextrin transporter activity. The combined mutations D433G and C82S increased the activity by 62%, while the single mutation C82S increased the transport rate by 46%. This demonstrates that the C82S mutation is responsible for nearly all observed increase in the transport rate (Table 2.3).

2.2.4 Structure/Sequence-Function Relationships

A homology model was developed for the β-glucosidase and the mutations were investigated for structure-function relationships. The homology model predicted that the L173H mutation was

located within the active site in the previously identified substrate entrance region (16). The wild-type L173 did not have any predicted interactions with cellobiose. While when the residue is switched to H173, it is predicted to have direct hydrogen-bonds to the hydroxyl group of the C1 atom (Figure 2.8). The H23L mutation is on the periphery of the enzyme, far from the active site (Figure 2.9).

No crystal structure is available for the cellodextrin transporter, therefore to visualize and hypothesize how these mutations affect the activity, a sequence-based analysis was performed with the aid of HMMTOP software (26,27). This software is used to predict transmembrane helix domains, helix tail domains, and extra/intracellular loop structures (Figure 2.10). Based on the predicted structure, the cellodextrin transporter is comprised of 12 transmembrane helices with one large inside loop and one large outside loop. The D433G mutation is predicted to be located on the large outside loop and the C82S mutations is predicted to be in the first transmembrane helix.

2.3 Discussion

Efficient production of biofuels and specialty chemicals hinges on the optimization of the metabolic pathways associated with the product. To this end, a variety of successful pathway optimization methods, mainly based on transcriptional engineering have been developed (14,28-34). Strategies in protein engineering are applied to overcome inefficiencies based on the innate characteristics and poor function of the enzymes themselves. Often, these enzymes are removed from the pathway context and engineered *in vitro* or *in vivo* for the highest activity. This task is arduously inefficient and there is no guarantee the engineered enzyme will reduce the bottleneck when it is reinserted into the pathway. Protein engineering strategies can now be applied to

multiple enzymes on the pathway scale due to recent advances in large-scale library creation (14,24,35,36). These library creation methods range from *in vitro* or *in vivo* assembly of plasmids or directly into the chromosome. Our study illustrates *in vivo* library creation of plasmids for the simultaneous engineering of multiple proteins in a pathway through directed evolution in *S.cerevisiae*. There has been only one example of the directed evolution of multiple proteins, which was illustrated in *Escherichia coli* for arsenate resistance (37). Our work optimizes a sugar utilization pathway with increased metabolite products with a simple one-step assembly method, allowing for quick and efficient rounds of evolution, representing the first example of pathway optimization through multiple protein engineering for increased biofuel production in *S.cerevisiae*.

The cellobiose utilization pathway offers different applications in biofuel production. This pathway was originally investigated for co-utilization with xylose to reduce the diauxic shift or glucose repression during pentose sugar fermentations (11-13). Perhaps more significantly, this cellobiose utilization pathway can reduce *in vitro* enzymes used in hydrolysis of the biomass. In current processes, the insufficient β-glucosidase activity results in an accumulation of cellobiose in the cellulose hydrolysate (38). Thus engineering cellobiose utilization into *S.cerevisiae* for fermentation to biofuels is significant. The hydrolytic and phosphorolytic pathways offer more efficient methods for utilization; however there is still an opportunity for optimization. Recent studies adapted the phosphorolytic pathway for improved efficiency (10), including some work to engineer the cellodextrin transporter through adaptive evolution and alanine scanning. The previous work in our laboratory optimized the hydrolytic pathway based on combinatorial transcriptional engineering, but we endeavored to engineer the proteins to improve the inherent characteristics of the enzymes.

The final optimized pathway (R2) increased the cellobiose consumption rate by 54% and ethanol productivity by 74%. Glucose and cellodextrin accumulation in the supernatant, which was previously observed in the hydrolytic pathway (13,17), was also noted in the R2 pathway. The accumulation occurred after all the cellobiose was utilized and was present only at 10 g/L. Application of this pathway for reduced *in vitro* enzyme pretreatment deems this by-product negligible. Highest specific growth rates on cellobiose are observed only when the β-glucosidase L173H mutation is present. The H23L mutation by itself did not lend a higher growth rate, even though it was coupled with the most active cellodextrin transporter. In the cellodextrin transporter, the most significant mutation was clearly the C82S mutation, as this mutation conferred the highest specific growth rates. However, both the β-glucosidase L173H and the cellodextrin transporter C82S mutations need to be present in order to achieve the highest growth rate. This synergistic effect demonstrates the power of simultaneously engineering multiple proteins.

The β-glucosidase, from glycoside hydrolase family 1 (GH1), contains the standard (α/β)-barrel structure of the enzyme family. The active site is located in a deep cleft formed by the connection loops at the C-terminal end of the β-sheets of the TIM barrel (17) and negatively charged residues surround the bottom of the active site. The conserved catalytic residues of the GH1 family involve the catalytic acid/base Glu 166 and the nucleophile Glu 377, which cleave the glycosidic bond via the common double displacement mechanism (15). This mechanism acts on the β-1,4-linked glucose derivatives, thus the enzyme is promiscuous with a range of substrates with varied affinities (17). The wild-type β-glucosidase enzyme has a 2.9-fold innate preference for *p*-NPG over cellobiose as a substrate. The directed evolution process performed in this study identified mutations in the enzyme that shift the specificity towards cellobiose. The

L173H mutation is shown to be the most critical mutation for substrate specificity. Homology modeling indicated the H173 residue can hydrogen-bond directly to the hydroxyl group of the C1 atom (Figure 2.8). In wild-type β-glucosidase, substrate docking is hypothesized to occur via an extensive network of polar interactions, where well-ordered water molecules and enzyme-substrate hydrogen bonding keep the cellobiose firmly in place during hydrolysis (15,18). Recently, there has been significant interest in the β-glucosidase GH1 family focusing on substrate specificity (15-18). This work has identified a residue that can further tailor the specificity of β-glucosidase towards cellobiose.

The H23L mutation is considered to be an overall activity enhancer. The single H23L mutation slightly improves overall enzyme activity towards both the *p*-NPG and cellobiose. H23 is located on the periphery of the enzyme, quite distant from the active site (Figure 2.9). It is probable that this mutation contributes to increased solubility or stability. As this amino acid residue was mutated from a hydrophilic to a hydrophobic side chain, it is possible the increased hydrophobicity allowed for more stable bonds between the peripheral residues.

The screening method operated at saturating cellobiose concentration at 80 g/L. As such, through this method the transporters were selected for increased V_{max}, as the K_M for the transporters were previously reported at 3 μM (11). In the study, the mutants obtained did exhibit increased specific activity, associated with an increased V_{max} of the transporter. Though this is a biased screen, any industrial process for cellobiose utilization will likely occur at saturated cellobiose conditions, making it the most relevant condition to select at. Recent studies in transporter engineering, which used a growth-based screen, also discovered mainly increases in V_{max} (10,19) rather than improved K_M. The D433G mutation increased activity over the wild-type by 10%. However, the combination of D433G+ C82S mutation increased the activity over

the wild-type by 62%. It was shown that this increase in activity was attributed mainly to the C82S change as the single mutation also had a 46% increase in activity over the wild-type. This was supported by the previous growth curve data wherein the highest growth rate occurred only when the C82S mutation was present.

Modeling of transporter enzymes to identify the structure-function relationships is difficult due to the complexity of crystallizing membrane proteins. Therefore, a prediction of the sequence-structure for the cellodextrin transporter was made using HMMTOP software (26,27). We can hypothesize and speculate how the mutations affected the phenotype by the location of the mutants on the predicted structure (Figure 2.10). The D433G mutation is predicted to be located on a large outside loop. As this mutation results in the smallest possible non-charged residue, it can be speculated that the outside loop causes a steric hindrance for the mass transfer of cellobiose to the membrane, which can be alleviated by the altered loop structure with the smaller non-reactive residue. The C82S mutation conferred the highest increase in activity and is predicted to be located in the first transmembrane helix. The location in the helix and the increased activity suggest that this mutation is directly associated with the protein complex stability or with the interaction with cellobiose. A recent report on adaptive evolution and alanine screening of the cellodextrin transporter identified mutants with increase V_{max}. The mutant with the highest activity in that study is also predicted to be associated with a membrane helix, while the other less improved mutants are predicted to be on the loop structures (10). A separate study in the directed evolution of xylose transporters identified mutations which tended to cluster around the first transmembrane helix (19), a similar finding to our study.

2.4 Conclusion

Cellobiose utilization has recently become a significant consideration in economical biofuel production. Efficient utilization will not only decrease the *in vitro* enzyme load in biomass pretreatment but it will also reduce the diauxic shift in pentose sugar utilization. We successfully optimized the pathway by simultaneously engineering multiple proteins in the pathway. This enabled the discovery of novel and synergistic mutations between the proteins, which conferred an improved phenotype of the entire pathway.

2.5 Materials and Methods

2.5.1 Strains, Media, and Culture Conditions

The industrial *Saccharomyces cerevisiae* strain Still Spirits (Classic) Turbo Distiller's Yeast was purchased from Homebrew Heaven (Everett, WA). Yeast strains were cultivated in YP media (1% yeast extract, 2% peptone) with 2% glucose (YPD) or 2-8% cellobiose (YPC). YPC with 8% cellobiose was used in fermentation analysis while 2% cellobiose was used for plate screening. *S.cerevisiae* strains were cultured at 30 °C with orbital shaking at 250 rpm for aerobic growth or 100 rpm for oxygen limited conditions. As needed, 200 µg/mL G418 (KSE Scientific, Durham, NC) supplemented YPD for plasmid selection. *Escherichia coli* DH5α (Cell Media Facility, University of Illinois at Urbana-Champaign, Urbana, IL) was used for recombinant DNA manipulations. *E.coli* strains were cultured in Luria broth (LB) (Fischer Scientific, Pittsburgh, PA) at 37 °C and 250 rpm, supplemented with 50 µg/mL ampicillin. Yeast and bacterial strains were stored in 15% glycerol at -80 °C. All chemicals were purchased from Sigma Aldrich or Fisher Scientific.

2.5.2 Plasmid and Strain Construction

Restriction enzymes were purchased from New England Biolabs (Ipswich, MA). All cloning work was performed through yeast homologous recombination mediated by the DNA Assembler method (23). Yeast plasmids were isolated using Zymoprep Yeast Plasmid miniprep II kit (Zymo Research, Irvine, CA), then transformed into *E.coli* for isolation of high purity DNA. *E.coli* plasmids were isolated using Qiagen Spin Plasmid Mini-Prep Kit (Qiagen, Valencia, CA). PCR fragments were purified by QIAQuick Gel Purification Kit (Qiagen, Valencia, CA).

The β-glucosidase gene *gh1-1* (GenBank Accession number XM_951090) from *Neurospora crassa* was expressed using a PYK promoter and an ADH terminator. The cellobiose transporter gene *cdt-1 (*GenBank Accession number XM_958708) from *N. crassa* was expressed using a TEF promoter and a PGK terminator as previously constructed (13). To transfer the pathway to the pRS-KanMX plasmid, primers kanMX-PYKp-F and PGKt-kanMX-R (Table 2.4) were used to amplify the full cellobiose utilizing pathway. To facilitate the creation of a library of cellobiose utilizing pathways, a helper plasmid was constructed. The helper plasmid contained the PYK promoter and the PGK terminator, separated by a unique restriction enzyme recognition site *BamHI* for plasmid linearization. The helper plasmid will later be linearized and used as a backbone for the library creation. The PYK promoter was amplified using primers kanMX-PYKp-F and PYKp-BamHI-PGKt-R (Table 2.4). The PGK terminator was amplified using primers PYKp-BamHI-PGKt-F and PGKt-kanMX-R (Table 2.4).

2.5.3 Library Creation

The ADH terminator and TEF promoter were not subjected to random mutagenesis, thus this cassette was amplified using primers middle-BGL-ADHt-F and middle-TEFp-CDT-R (Table

2.4). This fragment was transformed with the error-prone PCR fragments and linear helper plasmid. The *gh1-1* and *cdt-1* genes were subjected to error prone PCR and error rates of 2-3 bp mutations per 1kb were used for library creation. The library was created according to the previously described method (14,24). The total library size obtained was 10^4. Ten colonies were randomly selected from the YPD+G418 plate and their plasmids were isolated and sequenced to confirm the diversity of the library.

2.5.4 Library Screening

The library was screened on YPC plates for large colony selection (14,24). After confirmation of the top 5 clones with the fastest growth rate, the plasmids were isolated and retransformed. This retransformation ensured that no adaptation occurred in the screening process and the increased activity is from mutations on the plasmid. After retransformation, the cells were seeded overnight in YPD+G418, washed twice with sterile water, and then inoculated into 50 mL YPC in 250 mL flask at 100 rpm and 30 °C (Figure 2.4). The cultures were sampled and analyzed for growth and also for metabolite production as described previously (14,24). After confirmation of improved mutant pathways, the plasmids were isolated and sequenced to identify the mutations within the genes in the pathway. After confirmation of the mutations, that pathway was used as a template for a second round of error-prone PCR. The library creation and screening process was repeated.

2.5.5 Fermentation Analysis

Seed cultures were inoculated from freshly streaked frozen stocks. YPD+G418 seed cultures were used for the single/double mutant fermentation tests. YPC seed cultures were used for final

mutant pathway analysis. The seed cultures were grown at 30 °C and 250 rpm overnight, then washed with sterile water twice before inoculating 50 mL YPC in 250 mL un-baffled flasks to an initial OD of 0.2. The cultures were incubated at 30 °C and 100 rpm for oxygen limited growth. The cultures were sampled and analyzed for growth and metabolite production.

2.5.6 Construction of Single Mutants

The single mutant genes were constructed through site-directed mutagenesis and the mega primer PCR method with primers listed in Table 2.4. After PCR amplification, the genes were transformed into the Classic strain along with the ADHt/TEFp cassette into a linearized pRS-KanMx helper plasmid. The final plasmid was purified and sequenced for confirmation.

2.5.7 β-Glucosidase Enzyme Activity Assays

Classic strains harboring an empty vector, the wild-type, and mutant pathways were grown to mid-exponential phase and washed three times with potassium phosphate buffer (pH 7). A final cell mass equivalent to an OD of 20 was harvested. Cell free extracts were prepared through YPER Extraction Reagent (Thermo Scientific, Rockport, IL). 125 µL of YPER was used to lyse the cells for 20 minutes at 25 °C, vigorously shaken at 700 rpm in a thermomixer (Eppendorf, Germany). After lysing, the cell membrane was pelleted for 10 minutes with 15,000 rpm at 4 °C. The protein concentration was determined via the BCA protein assay kit (Pierce, Rockford IL), following the standard manufacturer protocol.

The lysate was tested for *p*-NPG activity by 1 mM *p*-NPG in 100 mM potassium phosphate buffer at pH 7. The colorimetric change was monitored at 405 nm in a 96-well Biotech Synergy 2 plate-reader (Winooski, VT) for 30 minutes at 30°C. The change in absorbance was converted

to units (1 U = 1 μmol of *p*-NP/min) using Beer-Lambert law. The extinction coefficient was determined by a standard curve using *p*-NP at pH 7. For cellobiose-based enzymatic assay, the linear range of the β-glucosidase was determined. The reactions were carried out at 60 mM cellobiose in 100 mM potassium phosphate buffer at pH 7. After addition of lysate, the reaction was allowed to react for 15, 30, 45, and 60 minutes at 30 °C. The reaction was stopped by boiling at 100 °C for 10 minutes. The samples were then centrifuged at 15,000 rpm for 5 minutes before being stored on ice. The amount of glucose which had been produced in the allotted time frame was then measured using the D-glucose kit (R-Biopharm, Germany). Standard manufacturer"s instructions were followed. One unit (U) is defined as micromoles of glucose produced per minute. The specific activity was determined by normalizing the rate to the amount of protein in the assay.

2.5.8 CDT Activity Assay

The cellodextrin transporter was assayed using the oil-stop protocol previously reported (11,25) with few modifications. Cultures of the wild-type and mutant pathways were grown to an OD of 15-20, washed three times with ice-cold assay buffer (30 mM MES-NaOH + 50 mM ethanol), and then normalized to an OD of 20. 50 μL of cells were added to 50 μL of [3H]-cellobiose at 30 °C and layered over 100 μL of silicone oil (Sigma 85419), incubated for 10, 20, 40, and 80 seconds. The cells were then centrifuged through the oil at 15,000 rpm for 30 seconds. After ethanol/dry-ice bath, the cell pellets were solubilized in NaOH overnight. The amount of [3H]-cellobiose present in the cells was then quantified via a liquid scintillation counter. The amount of labeled cellobiose that was taken up by the cell was plotted against the time of reaction. One unit (U) is defined as the micromoles of cellobiose taken up by the cell per min. The dry cell

weight of the cells was determined and the rate of the cellobiose uptake was normalized by the gram cell cry weight (gdcw).

2.5.9 Homology Modeling

A homology model of the β-glucosidase enzyme was constructed to identify the structure-function relationships of the mutations discovered. The gene encoding the β-glucosidase from *Trichoderma reesei* (17) (PDB accession code 3AHY) afforded the highest homology to the *gh1-1* gene from *N. crassa* with 73% sequence identity and was used as a template for homology modeling. The structure model of the β-glucosidase from *N. crassa* was constructed using the modeling program Molecular Operating Environment (Chemical Computing Group, Montreal, Canada). After constructing the homology model, the substrate in the co-crystal structure of *Neotermes koshunensis* β-glucosidase (PDB accession code 3VIK) was docked into the model (16). The model was energy minimized before ligand interactions were investigated. To identify the effects of the mutations, the mutations were introduced to the model and energy minimized again before investigating the ligand interactions.

2.6 References

1. Wyman, C.E. (2007) What is (and is not) vital to advancing cellulosic ethanol. *Trends Biotechnol.*, **25**, 153-157.
2. Lynd, L.R., Weimer, P.J., van Zyl, W.H. and Pretorius, I.S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.*, **66**, 506-577.
3. Elkins, J.G., Raman, B. and Keller, M. (2010) Engineered microbial systems for enhanced conversion of lignocellulosic biomass. *Curr. Opin. Biotechnol.*, **21**, 657-662.
4. Nakamura, N., Yamada, R., Katahira, S., Tanaka, T., Fukuda, H. and Kondo, A. (2008) Effective xylose/cellobiose co-fermentation and ethanol production by xylose-assimilating *S. cerevisiae* via expression of β-glucosidase on its cell surface. *Enzyme and Microb. Technol.*, **43**, 233-236.
5. Guo, Z.P., Zhang, L., Ding, Z.Y., Gu, Z.H. and Shi, G.Y. (2011) Development of an industrial ethanol-producing yeast strain for efficient utilization of cellobiose. *Enzyme Microb. Technol.*, **49**, 105-112.
6. Saitoh, S., Hasunuma, T., Tanaka, T. and Kondo, A. (2010) Co-fermentation of cellobiose and xylose using β-glucosidase displaying diploid industrial yeast strain OC-2. *Appl. Microbiol. Biotechnol.*, **87**, 1975-1982.
7. Machida, M., Ohtsuki, I., Fukui, S. and Yamashita, I. (1988) Nucleotide sequences of *Saccharomycopsis fibuligera* genes for extracellular β-glucosidases as expressed in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **54**, 3147-3155.
8. Sadie, C.J., Rose, S.H., den Haan, R. and van Zyl, W.H. (2011) Co-expression of a cellobiose phosphorylase and lactose permease enables intracellular cellobiose utilisation by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.*, **90**, 1373-1380.
9. Aeling, K.A., Salmon, K.A., Laplaza, J.M., Li, L., Headman, J.R., Hutagalung, A. and Picataggio, S. (2012) Co-fermentation of xylose and cellobiose by an engineered *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.*, **39**, 1597-1604.
10. Ha, S.J., Galazka, J.M., Joong Oh, E., Kordic, V., Kim, H., Jin, Y.S. and Cate, J.H. (2012) Energetic benefits and rapid cellobiose fermentation by *Saccharomyces cerevisiae* expressing cellobiose phosphorylase and mutant cellodextrin transporters. *Metab. Eng*.
11. Galazka, J.M., Tian, C., Beeson, W.T., Martinez, B., Glass, N.L. and Cate, J.H.D. (2010) Cellodextrin transport in yeast for improved biofuel production. *Science*, **330**.
12. Ha, S.J., Galazka, J.M., Kim, S.R., Choi, J.H., Yang, X., Seo, J.H., Glass, N.L., Cate, J.H. and Jin, Y.S. (2011) Engineered Saccharomyces cerevisiae capable of simultaneous cellobiose and xylose fermentation. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 504-509.
13. Li, S., Du, J., Sun, J., Galazka, J.M., Glass, N.L., Cate, J.H.D., Yang, X. and Zhao, H. (2010) Overcoming glucose repression in mixed sugar fermentation by co-expressing a cellobiose transporter and a β-glucosidase in *Saccharomyces cerevisiae*. *Mol. Biosyst.*, **6**.
14. Du, J., Yuan, Y., Si, T., Lian, J. and Zhao, H. (2012) Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Res.*, **40**, e142.
15. Isorna, P., Polaina, J., Latorre-Garcia, L., Canada, F.J., Gonzalez, B. and Sanz-Aparicio, J. (2007) Crystal structures of *Paenibacillus polymyxa* β-glucosidase B complexes reveal the molecular basis of substrate specificity and give new insights into the catalytic machinery of family I glycosidases. *J. Mol. Biol.*, **371**.

16. Lee, H.L., Chang, C.K., Jeng, W.Y., Wang, A.H. and Liang, P.H. (2012) Mutations in the substrate entrance region of β-glucosidase from *Trichoderma reesei* improve enzyme activity and thermostability. *Protein Eng. Des. Sel.*, **25**, 733-740.
17. Jeng, W.Y., Wang, N.C., Lin, M.H., Lin, C.T., Liaw, Y.C., Chang, W.J., Liu, C., Liang, P.H. and Wang, A.H.J. (2011) Structural and functional analysis of three β-glucosidases from bacterium *Clostridium cellulovorans*, fungus T*richoderma reesei* and termite *Neotermes koshunensis*. *J. Struct. Biol.*, **173**.
18. Nam, K.H., Sung, M.W. and Hwang, K.Y. (2010) Structural insights into the substrate recognition properties of β-glucosidase. *Biochem. Biophys. Res. Commun.*, **391**, 1131-1135.
19. Young, E.M., Comer, A.D., Huang, H. and Alper, H.S. (2012) A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. *Metab. Eng.*, **14**, 401-411.
20. Bokma, E., Koronakis, E., Lobedanz, S., Hughes, C. and Koronakis, V. (2006) Directed evolution of a bacterial efflux pump: adaptation of the *E. coli* TolC exit duct to the *Pseudomonas* MexAB translocase. *FEBS Lett.*, **580**, 5339-5343.
21. Dunlop, M.J., Dossani, Z.Y., Szmidt, H.L., Chu, H.C., Lee, T.S., Keasling, J.D., Hadi, M.Z. and Mukhopadhyay, A. (2011) Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.*, **7**, 487.
22. Ren, C., Chen, T., Zhang, J., Liang, L. and Lin, Z. (2009) An evolved xylose transporter from *Zymomonas mobilis* enhances sugar transport in *Escherichia coli*. *Microb. Cell. Fact.*, **8**, 66.
23. Shao, Z.Y., Zhao, H. and Zhao, H.M. (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.*, **37**.
24. Kim, B., Du, J., Eriksen, D.T. and Zhao, H. (2012) Combinatorial design of a highly efficient xylose utilizing pathway for cellulosic biofuels production in *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, ***In press***.
25. Arendt, C.S., Ri, K., Yates, P.A. and Ullman, B. (2007) Genetic selection for a highly functional cysteine-less membrane protein using site saturation mutagenesis. *Anal. Biochem.*, **365**, 185-193.
26. Tusnady, G.E. and Simon, I. (1998) Principles governing amino acid composition of integral membrane proteins: Application to topology prediction. *J. Mol. Biol.*, **283**.
27. Tusnady, G.E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics*, **17**.
28. Alper, H., Fischer, C., Nevoigt, E. and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 12678-12683.
29. Alper, H. and Stephanopoulos, G. (2007) Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab. Eng.*, **9**, 258-267.
30. Pfleger, B.F., Pitera, D.J., D Smolke, C. and Keasling, J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.*, **24**, 1027-1032.
31. Salis, H.M., Mirsky, E.A. and Voigt, C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946-950.
32. Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R. and Church, G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894-898.

33. Warnecke, T.E., Lynch, M.D., Karimpour-Fard, A., Lipscomb, M.L., Handke, P., Mills, T., Ramey, C.J., Hoang, T. and Gill, R.T. (2010) Rapid dissection of a complex phenotype through genomic-scale mapping of fitness altering genes. *Metab. Eng.*, **12**, 241-250.
34. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B.A. and Gill, R.T. (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.*, **28**, 856-U138.
35. Wingler, L.M. and Cornish, V.W. (2011) Reiterative Recombination for the *in vivo* assembly of libraries of multigene pathways. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 15135-15140.
36. Ramon, A. and Smith, H.O. (2011) Single-step linker-based combinatorial assembly of promoter and gene cassettes for pathway engineering. *Biotechnol. Lett.*, **33**, 549-555.
37. Crameri, A., Dawes, G., Rodriguez, E., Jr., Silver, S. and Stemmer, W.P. (1997) Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nat. Biotechnol.*, **15**, 436-438.
38. Bommarius, A.S., Katona, A., Cheben, S.E., Patel, A.S., Ragauskas, A.J., Knudson, K. and Pu, Y. (2008) Cellulase kinetics as a function of cellulose pretreatment. *Metab. Eng.*, **10**, 370-381.

2.7 Tables

Table 2.1: Parameters for the improved pathways identified through directed evolution. Key parameters for an improved pathway include specific growth rate, cellobiose consumption, and ethanol productivity. Errors are derived from biological duplicates. Fermentations conditions in 8% cellobiose rich media and oxygen limited conditions.

	Wild-type	**R1**	**R2**
Specific Growth Rate (h^{-1})	0.053 ± 0.002	0.063 ± 0.0003	0.075 ± 0.0002
Cellobiose consumption (g cellobiose/(L·h))	1.68 ± 0.09	2.35 ± 0.07	2.815 ± 0.15
Ethanol productivity (g ethanol/(L·h))	0.59 ± 0.02	0.674 ±0.03	1.03 ± 0.03
Yield (g ethanol/g cellobiose)	0.418 ± 0.002	0.437 ± 0.002	0.447 ± 0.004

Table 2.2: Specific activity measurements from crude lysate of the engineered β-glucosidase isolated directly from the measured strain. One unit is defined as a μmol/min. Each mutant was quantified with *p*-NPG and cellobiose as a substrate to demonstrate the increased substrate specificity towards cellobiose.

	p-NGP (U/mg protein)	***Cellobiose (U/mg protein)***
Empty Vector	0.0072 ± 0.0001	*ND**
Wild-type	0.194 ± 0.011	0.066 ± 0.0002
L173H	0.133 ± 0.177	0.123 ± 0.003
L173H + H23L	0.156 ± 0.009	0.163 ± 0.006
H23L	0.219 ± 0.006	0.078 ± 0.003

ND* = not detected

Table 2.3: Specific activity measurements from the cellodextrin transporter as determined through radioactive-labeled cellobiose uprate assay. One unit is defined as 1 µmol/min.

	Cellobiose Uptake Rate (*U/gcdw*)
Empty Vector	0.006 ± 0.000
Wild-type	1.760 ± 0.109
D433G	2.084 ± 0.134
D433G+C82S	2.849 ± 0.358
C82S	2.565 ± 0.093

Table 2.4: Primers used in this study

PGKt-kanMX-R	5″-CCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCAGGAAGAATACACTATAC-3″
BGL EP-F	5″-CTCTCTTGTTTCTATTTACAAGACACCAATCAAAACAAATAAAACATCATCACAATG-3″
BGL EP-R	5″-GCCGACAACCTTGATTGGAGACTTGACCAAACCTCTGGCGAAGAAGTCCAAAGCTTTA-3″
CDT EP- F	5″-CTTCTTGCTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTAATTACAAAATG-3″
CDT EP-R	5″-GGATGGGGAAAGAGAAAAGAAAAAAATTGATCTATCGATTTCAATTCAATTCAATCTA-3″
middle-BGL-ADHt-F	5″-GAAGCCGCTCTTTGACTCTTTGATCAAGAAGGACTAAAGCTTTGGACTTCTTCGCCAGAG-3″
middle-TEFp-CDT-R	5″-CGGTGCTGGCCCCGTCATGGGAGCCGTGAGACGACATTTTGTAATTAAAACTTAGATTAG-3″
B2-H23L-F	5″-GGGTGCTATCCTCGCCGACGGCCGTGGC-3″
B2-H23L-R	5″-GCCACGGCCGTCGGCGAGGATAGCACCC-3″
C2-C82S-F	5″-CGCCTTTTGTAGTGCATGCGCCAACGG-3″
C2-C82S-F	5″-CCGTTGGCGCATGCACTACAAAAGGCG-3″
qPCR-BGL-F	5″-AACCAGAAGGGCATCGACCACTAT-3″
qPCR-BGL-R	5″-GCACTTGGGAATGGCCTTGAACAT-3″
qPCR-CDT-F	5″-AATCCCCTCGCTTCCTATTTG-3″
qPCR-CDT-R	5″-TCCTGATACCGTCCCTCATC-3″
CDT sequencing-TEPp-F	5″-CTTTCGATGACCTCCCATTGATATTTAAG-3″
CDT sequencing- PGKt-R	5″-CGAATGGGAAAAAAAAACTGCATAAAGGC-3″
BGL sequencing-F	5″-CCTTTGGTTTTTTATTCTTAACTTG-3″
BGL sequencing- R	5″-GAAGTGTCAACAACGTATCTACCAACG-3″

2.8 Figures

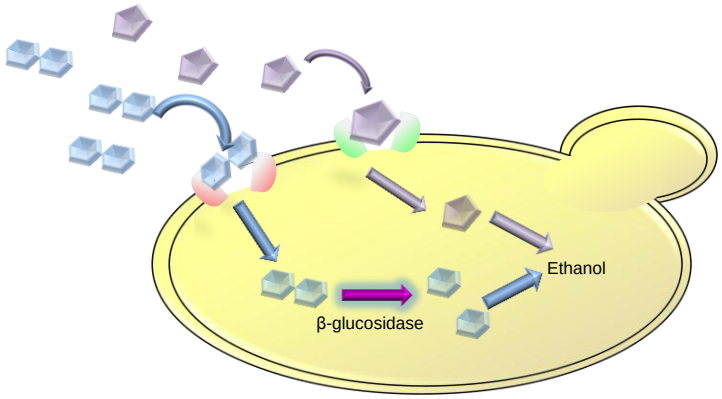


Figure 2.1: The hydrolytic cellobiose utilization pathway used in this study.

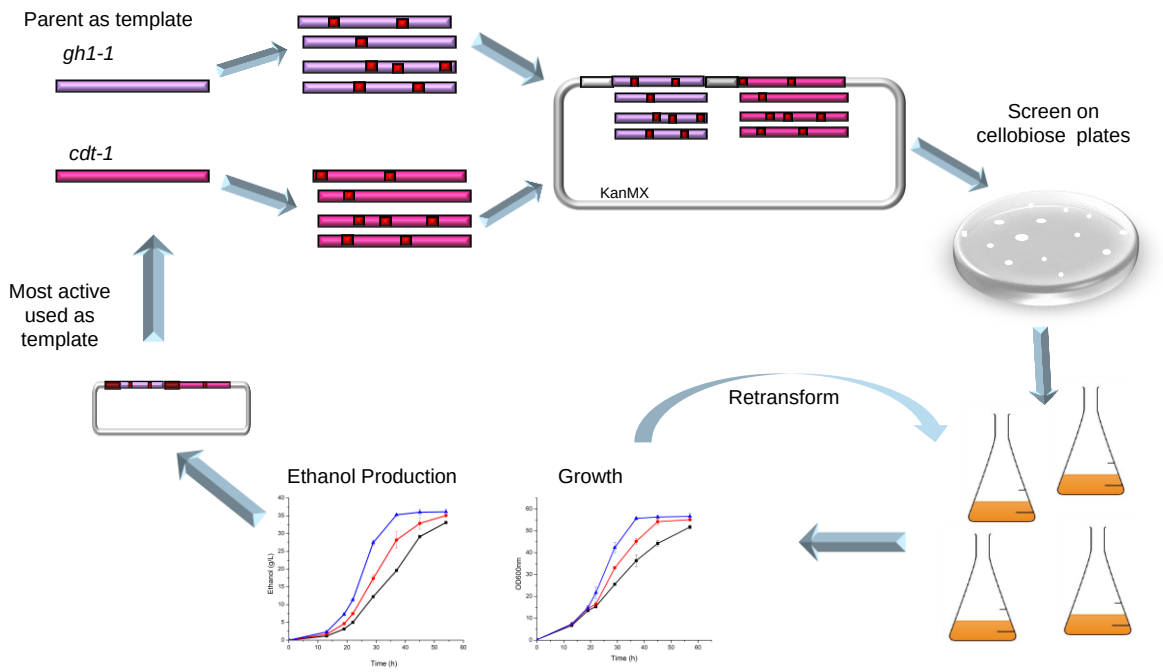
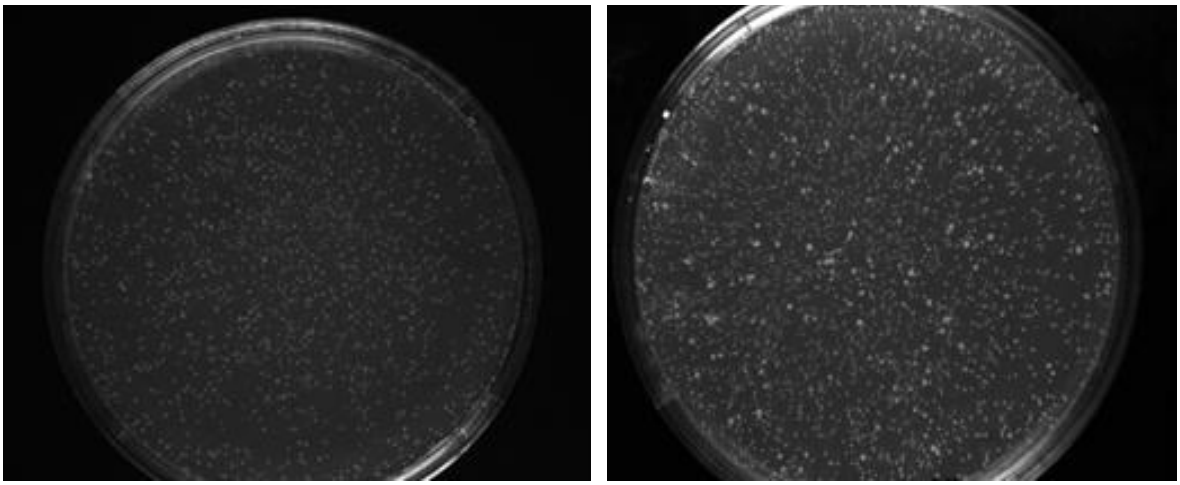


Figure 2.2: A general scheme for the directed evolution of multiple proteins within a metabolic pathway. Diversity is introduced to the genes through error-prone PCR, DNA shuffling, or other mutagenesis techniques. The genes are then recombined through homologous recombination into the pathway. A non-mutagenized fragment for the interior - promoter and terminator is also amplified and assembled. Screening is accomplished via colony size on cellobiose agarose plates. Large colonies are associated with faster growth and selected for further analysis. To confirm phenotype is a result of the improved pathway and not from adaption, the pathways are retransformed. Improved mutants are selected for a second round of diversification and screening.



(a) (b)

Figure 2.3: Selection plates of the error-prone library on cellobiose plates, depicting the large colonies. (a) wild-type pathway with all uniform colonies (b) error-prone PCR library.

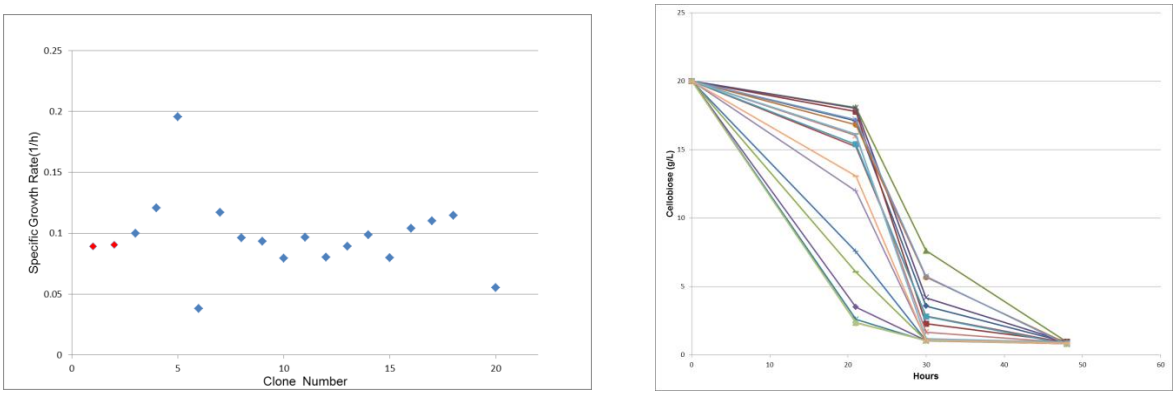
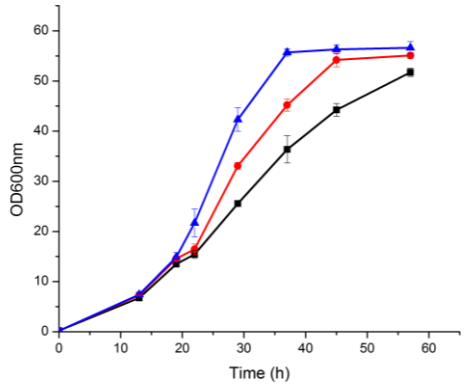
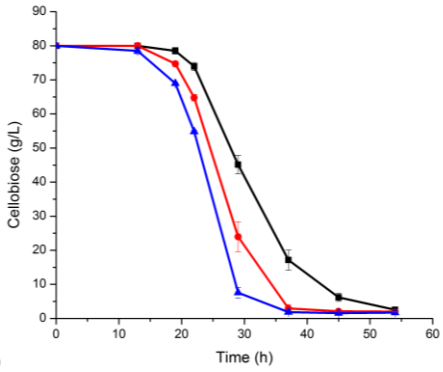


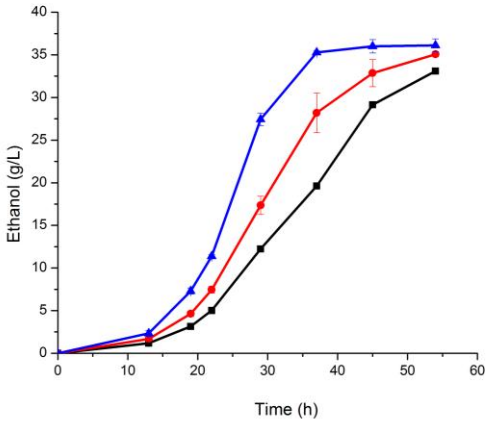
Figure 2.4: (a) Specific growth rates of mutants in the top 10% fastest growers in the initial tube screening of the library created by the first round of error-prone PCR. (♦) is the wild-type pathway (♦) depicts the selected clones, numbered on the x-axis. (b) Cellobiose utilization from confirmation of clones with high specific growth rate fermentation carried out in oxygen limited conditions in un-baffled flasks. (■) (♦) denote the wild-type pathway cellobiose consumption, (■) (x) represent two of the faster cellobiose utilizing pathways.



(a)



(b)



(c)

Figure 2.5: Fermentation profile of the cellobiose utilization pathways under oxygen limited conditions: (■) wild-type pathway, (●) R1 pathway, (▲) R2 pathway (a) Growth curve, (b) Cellobiose utilization, (c) Ethanol production. Error bars derived from biological duplicates.

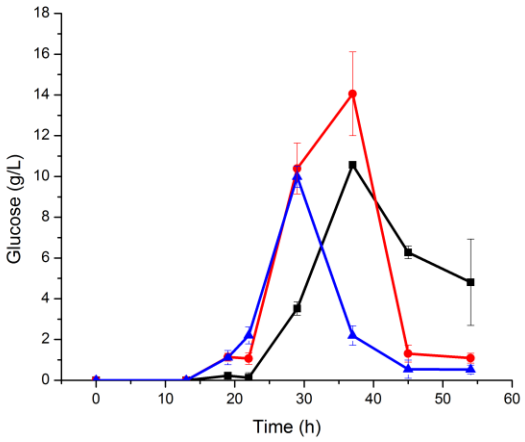
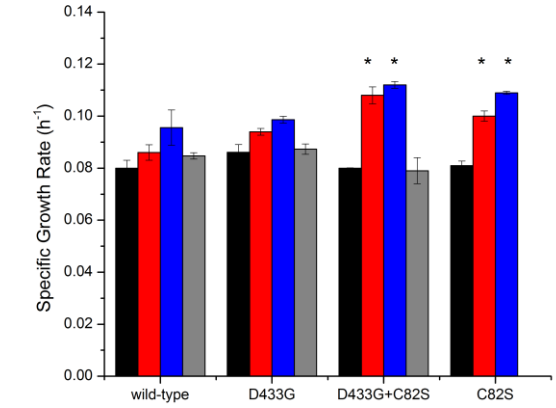
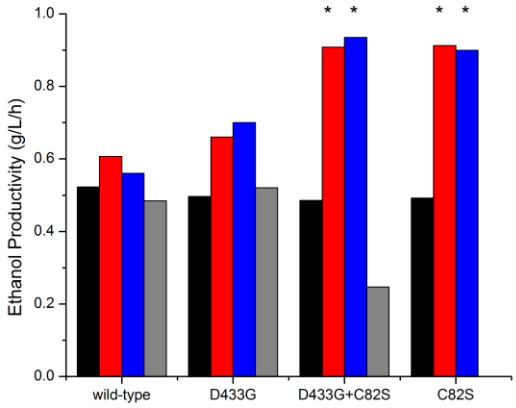


Figure 2.6: Glucose accumulation in oxygen limited fermentation (■) wild-type pathway, (●) R1 pathway, (▲) R2 pathway.

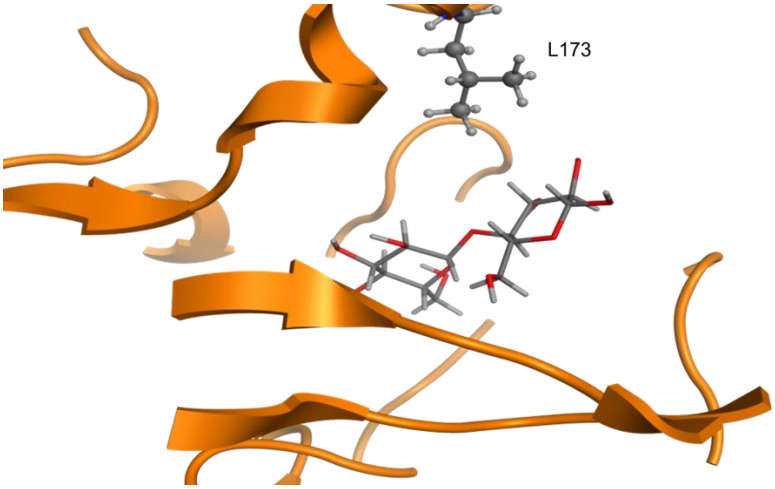


(a)

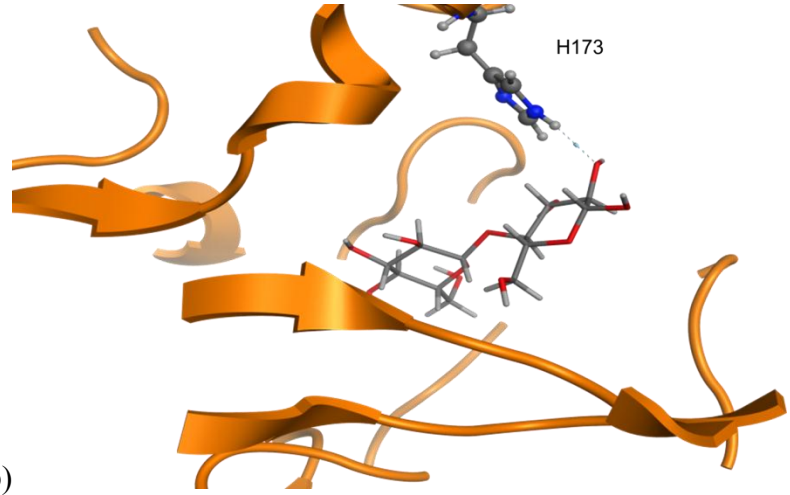


(b)

Figure 2.7: (a) Growth rate and (b) ethanol productivity of the single and double mutants on cellobiose as a sole carbon source. Error bars are derived from biological duplicates. Each bar represents the mutation in the β-glucosidase coupled with that of the cellodextrin transporter. The black bar is the wild-type, the red bar is the L173H, the blue bar represents the double mutant L173H and H23L, and the final grey bar is the single H23L. * denote statistically significant values over the wild-type ($p < 0.05$). Error bars are currently being completed for (b).



(a)



(b)

Figure 2.8: Homology model of the *N. crassa* β-glucosidase with the cellobiose substrate docked into the active site. (a) The critical residue in the wild-type enzyme L173 has no predicted interaction with the cellobiose molecule. (b) The mutated residue H173 is predicted to have direct hydrogen bonding with the hydroxyl group on the C1 atom of the cellobiose molecule.

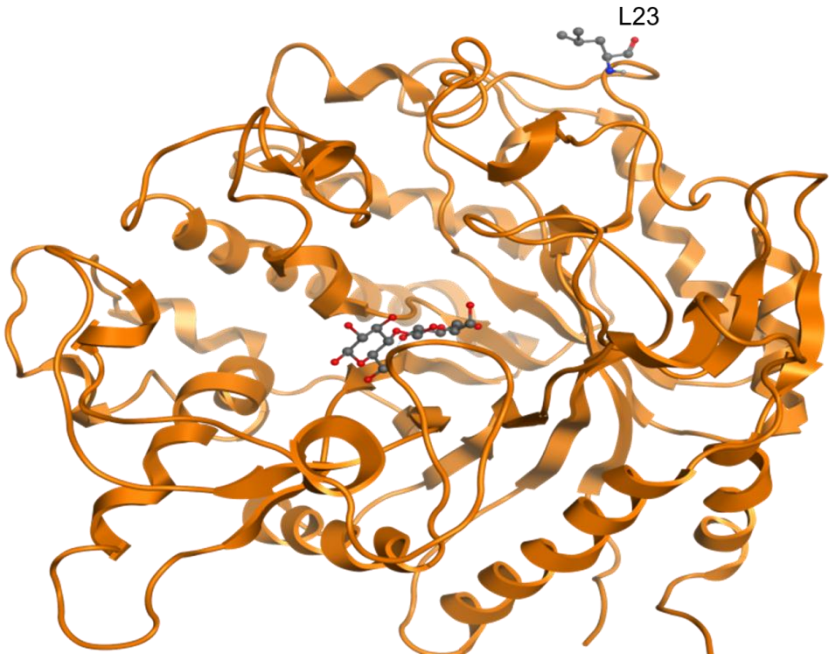


Figure 2.9: Homology model of the *N. crassa* β-glucosidase with the cellobiose substrate docked into the active site. The L23 mutation is notably far from the active site and the cellobiose substrate down in the center.

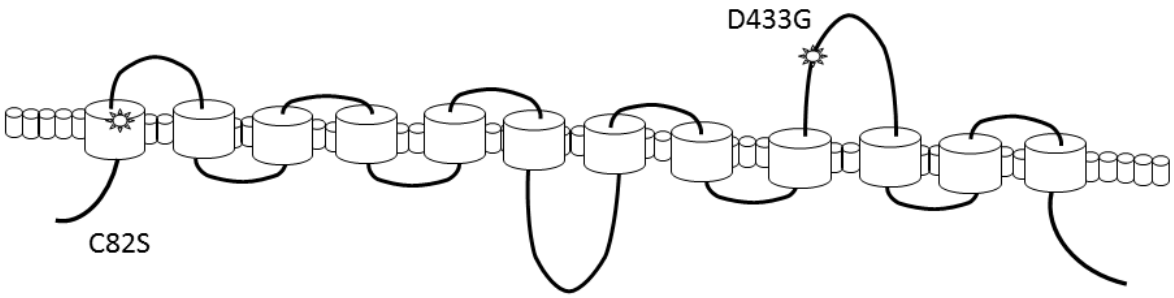


Figure 2.10: Sequence-structure based mapping of the cellodextrin mutations. The HMMTOP software predicted sequence-structure of the transmembrane helix, inner, and outer loops of the cellodextrin transporter in the cell membrane. The top of the graphic depicts outside of the cell, while the bottom. The mutations were overlaid onto the structure prediction.

Chapter 3: Engineering a Novel Pathway for FAEE Biosynthesis

3.1 Background

Industrial-scale bioethanol production through yeast has been marginally successful; however it is not a preferred biofuel molecule. Bioethanol has low energy intensity, high vapor pressure and is corrosive to engines. Thus, there is a great interest in microbial production of advanced biofuels such as butanol, propanol, alkane, alkenes, and biodiesel.

Biodiesel is a particularly interesting fuel molecule, as it can be easily derived from fatty acids. Fatty acids are only one catalytic step away from petroleum-derived diesel molecules and thus have become a focus in advanced biofuel production. Fatty acid ethyl esters (FAEEs), another name for biodiesel, are a non-toxic and completely biodegradable fuel. Compared to gasoline, using FAEEs can reduce emissions due to less carbon monoxide, sulfur, aromatic hydrocarbons and soot particles produced from the combustion process (1). FAEEs were previously produced via a transesterification reaction from plant or animal oils with ethanol. This process is energy intensive and expensive, thus recent microbial engineering efforts have focused on producing FAEEs. Figure 3.1 shows a simplified general scheme for the microbial production of FAEEs.

3.1.2 Fatty Acid Synthesis

FAEEs can be produced microbially from fatty acids. Therefore, maximum production of fatty acids has become a focus in recent research projects. Though there are many different platforms and systems to produce fatty acids in different organisms, the overall catalytic mechanisms are the same. The first committed step in fatty acid biosynthesis is the carboxylation of acetyl-CoA by acetyl-CoA carboxylase (ACC). This forms malonyl-CoA, which is subsequently converted by malonyl CoA:ACP transacylase (FabD) to malonyl-ACP. A series of iterative chain

elongation steps then occur in a cyclical fashion. The growing acyl chain is covalently attached to the acyl-carrier protein (ACP) through a thioester linkage on the terminal sulfhydryl of the phosphopantetheine prosthetic group. The first elongation step is initiated by the Claisen condensation of malonyl-ACP with an acyl-CoA, catalyzed by the first condensing enzyme β-ketoacyl-ACP synthase III (FabH). This forms β-ketoacyl-ACP. The FabR-FabB proteins (3-ketoacyl-ACP synthases) condense malonyl-ACP with acyl-ACP to extend the acyl chain by two carbons. This is the only irreversible step in the elongation cycle. The NADPH-dependent β-ketoacyl-ACP reductase FabG reduces the β-keto group to a β-hydroxy intermediate. This intermediate is then dehydrated by the β-hydroxyacyl-ACP dehydratase known as FabZ or FabA to a *trans*-2-enoyl-ACP. The final step reduces the enoyl chain by the NADPH-dependent FabI, FabL, or NADH dependent FadV enoyl-ACP reductase. After each complete cycle of condensation and chain reduction, the growing acyl chain is transferred back to the ketosynthase to initiate the next cycle (2) (Figure 3.2).

These reactions are catalyzed by two classes of enzymes which are divided by their architecture. Type II fatty acids are comprised of distinct individualized enzymes which independently complete the condensations and reductions. These systems are generally found in bacteria and other prokaryotes. Eukaryotes generally have a Type I fatty acid synthase (FAS) (Figure 3.3), which is a large globular enzyme. The microbial FAS enzyme is a hexamer protein with the domain arrangement of acyltransferase (AC), enoyl reductase (ER), dehydratase (DH), malonyl/palmitoyl transferase (MPT), acyl carrier protein (ACP), 3-ketoacyl reductase (KR), and 3-ketoacyl synthase (KS). The FAS of yeast and mycobacteria use an integral MPT activity for transacylation of the acyl-enzyme product. The fatty acid biosynthesis is initiated by the FAS component enzyme acetyltransferase, loading the acyl primer, usually acetate, from coenzyme A

(CoA) to a specific binding site on the FAS (3). Termination occurs by removing the product from FAS either by transesterification to an appropriate acceptor, usually palmitoyl transferase or a thioesterase. The iterative chain elongation mechanism involves enzyme bound intermediates being cycled through the catalytic domains. The chain length of the product is an inherent characteristic of each individual FAS (3). Bacterial type I fatty acid synthases have shown to be different from other FASs by being primarily bimodal in the chain length products, generally producing a chain length of 16 to 18 and also 24 to 30 carbons. Bacterial type I fatty acid synthases also differ from yeast FASs by utilizing the cofactors NADPH and NADH (3). Additionally, the bacterial apoFAS activating phosphopantetheine transferase is encoded by a separate gene, rather than being integrated into the large FAS, which occurs in yeast.

This research focuses on the use of a heterologously expressed bacterial FAS gene for synthesis of FAEEs in *S.cerevisiae*. This FAS gene was isolated from the bacteria *Brevibacterium ammoniagenes* (4,5) and was previously shown in our laboratory to be active in *S.cerevisiae* (6). Unlike other FAS enzymes, this particular FAS is unique because it does not produce fatty acids at varying chain lengths, it produces palmitic acid, C_{16}, almost exclusively with no unsaturated fatty acids produced (5). Having a uniform product will be significant for future biofuel applications, to allow for a consistent product. In this project, the FAS is co-expressed with its phosphopantetheine transferase PPT1, as this domain is not a part of the larger complex. The FAS is a very large enzyme, at 9.5 kb, which offers problems when cloning. Further details on the specifics of this research project are explained below.

3.1.2.1 Overproduction of Fatty Acids and Biofuel Production

Significant research has recently been accomplished producing biofuels from fatty acids. One example is the n-butanol production from the reversal of β-oxidation (7) (Figure 3.4). During the β-oxidation process fatty acids are broken down in the mitochondria to generate acetyl-CoA. By expressing the following enzymes the actual β-oxidation process is reversed: the thiolase FadA, then a hydroxyacyl-CoA dehydrogenase FadB, also the enoyl-CoA hydratase FadB, and finally an enoyl-CoA reductase YdiO. The reversal will actually build up fatty acids, which can be used to produce butanol products after the over-expression of an alcohol dehydrogenase and an acyl-CoA thioesterase. As a proof of concept, the product of one reverse cycle, n-butanol, was produced at 12 g/L from this pathway.

Another example of increased fatty acid production was from the Pfleger group (8) (Figure 3.5) . In this project, an *E.coli* strain that overproduces medium-chain length fatty acids via three basic modifications was accomplished. These modifications included elimination of β-oxidation through gene knockout of *fadD*, overexpression of the four subunits of acetyl-CoA carboxylase, and the expression of a plant acyl-acyl carrier protein (ACP) thioesterase. The thioesterase was the most significant component of this work, as this is the enzyme which pulled the fatty acid intermediates from the cyclical elongation process. The excess fatty acids were extracted and then subsequently converted to alkanes. A similar effort was also considered by Steen *et al.* where the thioesterase was expressed in the cytoplasm, instead of the periplasm, the truncated „TesA allowed for increased production of free fatty acids (9).

In a final example, Kalscheuer *et al.* overexpressed an unspecific acyltransferase from *Acinetobacter baylyi* (1), which produced microdiesel in *E.coli*. This unspecific acyltransferase catalyzes the esterification of a fatty acyl-CoA and a fatty alcohol to produce the FAEE. This

method was only possible by overexpression of a pyruvate decarboxylase and an alcohol dehydrogenase. Thus, the ethanol formation from the pyruvate decarboxylase and an alcohol dehydrogenase was subsequently esterified by the acyltransferase with the acyl moieties of coenzyme A thioesters of fatty acids to produce FAEEs. 1.25 g/L of FAEEs were produced after 70 hours using this method (1). Further information about the acyltransferase is available in the next section.

3.1.3 Wax Ester Synthase/acyl-coenzyme A:Diacylglycerol Acyltransferase

The wax ester synthase/acyl-coenzyme A:diacylglycerol acyltransferase (WSDGAT) enzyme is a relatively new family of acyltransferase enzymes. This enzyme exhibits bifunctional catalytic activities (10,11) (Figure 3.6) and is responsible for catalyzing the esterification of a fatty acyl-CoA and a fatty alcohol. This class of enzymes is instrumental in energy storage and lipid accumulation. It was first discovered in *Acinetobacter baylyi* (1) and a few other homologs have been cloned (12,13). Sequence identity between the homologous enzymes is about 20%, which is considered to be low. However, the enzymes share the assumed active site motif (HHXXXDG) (14). Recent studies have investigated the composition of the natural wax esters or TAGs produced by these enzymes when they are heterologously expressed in *E.coli* to determine the substrate specificities of the enzymes (14). The lipid profiles produced indicate a large variation in the types of neutral lipids they accumulate, thus confirming the promiscuity of the enzyme. A second study noted that the wax ester synthases naturally accept acyl groups with carbon chain lengths of C_{16} or C_{18} and linear alcohols with carbon chain lengths ranging from C_{12} to C_{20} (15). In follow-up work, a sequenced-based homolog study identified residues affecting the fatty alcohol selectivity (16). In this work, the mutation in the protein Ma1 from

Marinobacter aquaeolei VT8, A360I, was originally investigated to block the putative substrate pocket from large fatty acids. While this mutation did not affect large fatty acid binding, it was found to improve the rate of reaction with smaller fatty acids such as nonanol. The mutant exhibited an improvement in specific activity that was 8-fold greater than the wild-type. Further mutations to the site did not block large fatty acid binding. This result was also found in the previously characterized WSDGAT from *A. baylyi*. As the crystal structure is not yet available for this enzyme, further discussion about the structure-function relationship of this mutation cannot occur.

The most interesting of these recent studies comes from the functional expression and characterization of five wax ester synthases in *S.cerevisiae* and their utility for biodiesel production (15). In this study, five WSDGATs were examined in crude cell extracts of *S.cerevisiae.* The wax ester synthases (WSs) were from *Acinetobacter baylyi, Marinobacter hydrocarbonoclasticus, Rhodococcus opacus, Mus musculus,* and *Psychrobacter articus*. It was shown that the WS2 from *M. hydrocarbonoclasticus* was the most efficient at biodiesel production in *S.cerevisiae*. This enzyme produced 6.3 mg/L of FAEEs. No FAEEs were detected in the control strain. In this study, the acetyl coenzyme A carboxylase was also over-expressed, which resulted in a 30% increase in biodiesel production across all strains. Based on the recent success of the study, the WS2 was introduced into the experiments of this thesis.

3.1.4 Project Goal

For this project, it was desired to overproduce fatty acids via a bacterial FAS and then have the fatty acids subsequently converted to FAEEs by a WSDGAT in *S.cerevisiae* (Figure 3.7). A major limitation in fatty acid overproduction is the complex endogenous regulatory network that

limits fatty acid synthesis. Therefore, we sought to overproduce fatty acids through the heterologous expression of a bacterial FAS: it is expected that the bacterial FAS will not be subjected to the endogenous regulation of fatty acid synthesis in yeast. Thus, the production of fatty acids will not be limited by the mechanistic regulation. Heterologous expression of FAS for overproduction of fatty acids has not previously been accomplished; this is the main novelty and significance of this project. The bacterial FAS we chose to overexpress is the *Brevibacterium ammoniagenes* FAS-B because this enzyme exclusively produces a single-length fatty acid: C_{16}. Also, the FAS-B was previously proven to have functional expression in *S.cerevisiae* (6). The WSDGATs were used based on their relatively high activity shown in the literature (13,15,17). The chosen chassis, *S.cerevisiae*, is significant for its endogenous production of ethanol. Other studies which produced FAEEs in *E.coli* required the introduction of the ethanol pathway (1,9). Shi *et al.* recently illustrated the advantage of using the endogenous ethanol pathway in *S.cerevisiae* (15). Yeast is also more desirable than *E.coli* as an industrial host due to its robustness, phage resistance, and ability to grow at low pH, which reduces the risk of contamination.

3.2 Results

3.2.1 Plasmid Construction

Three WSDGAT genes have been cloned, including *WS2* [GenBank Accession No: ABO21021.1], *mhWSDGAT* [GenBank Accession No: ABO21020.1], and *abWSDGAT* [NCBI Reference Sequence: YP_045555.1]. The WS2 gene was codon optimized and synthesized by DNA 2.0 (Menlo Park, CA), while the other two WSDGAT genes from *Marinobacter hydrocarbonoclasticus* and *Acinetobacter baylyi* were cloned from the cDNA of the organisms.

The *fas-B* and *PPT1* were cloned from pGM44 (4). For ease of construction, the project was apportioned into two separate modules: module 1 involved expression of *fas-B* and *PPT1* while module 2 involved expression of the WSDGAT genes. In module 1, the plasmids included both a high and low copy version of: pRS424-TPIp-FAS-TPIt-TEFp-PPT1-TEFt. In module 2, there were three plasmids, in high and low copy versions of: pRS425-PYKp-mhWSDGAT-PYKt, pRS425-PYKp-abWSDGAT-PYKt, pRS425-PYKp-WS2WSDGAT-PYKt (Figure 3.8). The plasmids were first assembled into the low copy plasmids before being transferred into the high copy plasmid as a single gene cassette. The high copy plasmids were used for all further characterization. Construction of these plasmids was confirmed via PCR fragments, DNA restriction digests, and sequencing. By separating the pathway into modules with individual expression of the enzymes, separate expression, and activity assays could be performed. However, the modules are currently being combined for biodiesel production. A fourth gene was included in the full pathway, an acyl-CoA ligase, to convert the fatty acids to fatty acyl-CoA. The full biodiesel pathway and combination of the modules will result in three plasmids. For simplicity, the promoter/terminator names are not included here, but are the same as in the module stated above: pRS424- FAS-PPT1-FAA1-mh WSDGAT, the second plasmid is FAS-PPT1-FAA1-ab WSDGAT, and the third plasmids is FAS-PPT1-FAA1-WS2 WSDGAT (Figure 3.9).

3.2.2 Expression Studies

The *FAS-B* and *WSDGATs* were tested for expression in both *E.coli* and *S.cerevisiae* through SDS-PAGE experiments and Western blot analysis. SDS-PAGE experiments in *E.coli* for the abWSDGAT and mhWSDGAT showed positive expression. A clear band at about 55 kDa was

shown for ab- and mh- WSDGAT in *E.coli* cultures (Figure 3.10). The FAS-B expression in *E.coli* was not as easily identified, even after efforts to concentrate the sample significantly, the FAS was barely visible on the SDS-PAGE gel. The figure is not shown as the band was so faint, it was not visible in a photographic image.

The Western blots analysis for expression in *S.cerevisiae* was not successful. In these experiments, after efforts to significantly concentrate the sample, a faint band for the mhWSDGAT could be seen. The figure is not shown as the band was so faint, it was barely visible in a photographic image. However no bands for the abWSDGAT or the FAS-B were visible. The WS2 WSDGAT was not tested because at this time, we had not yet identified the advantage of the WS2 enzyme.

3.2.3 qPCR

To further test the expression of the WSDGAT enzymes and the FAS-B, a more sensitive method was used. Quantitative-PCR was used to detect minute amounts of the mRNA from these enzymes. The FAS-B and PPT1 were tested and shown to have significant level of mRNA compared to the negative control. The abWSDGAT was also tested, and shown to have significant levels of mRNA compared to the negative control. The mhWSDGAT was not tested because construction of the plasmid was not completed at the time. The WS2 was not subjected to the qPCR at this time because it had not yet been identified for use. All expression levels are normalized to the housekeeping gene asparagine-linked glycosylation 9 gene (*alg9*) expression to determine expression levels (Figure 3.11).

3.2.4 WSDGAT Assay

The activity assay for the WSDGAT measured the fatty alcohol and fatty acyl-CoA esterification catalyzed by the WSDGAT enzyme. This was done by monitoring the reaction of 5,5'-dithio-*bis*-(2)-nitrobenzoic acid (DTNB) with the free sulfhydryl-CoA released during the esterification reaction. The reaction of DTNB with sulfhydryl produces NTB^{2-}, a yellow-colored compound which can be quantified at a wavelength of 412 nm (Figure 3.12). This approach has been used by other research groups (13,14,18). It was found that the order in which substrate and enzymes are added in this assay is very important. For this particular work, the activity in the crude lysate of *S.cerevisiae* was being tested and ethanol was used as the co-substrate. This was a very noisy assay, wherein the negative control was very high. However, we have shown that both the mh and ab WSDGAT have activity significantly higher than the negative control in *S.cerevisiae*. At this time, the WS2 was not tested; however future work will repeat these assays with the WS2. One unit is defined as a µmole/min of NTB produced, which is stoichiometrically equivalent to a µmole/min of thiol being released. The activity was normalized by the total concentration of protein in the lysate. The background was subtracted from the final assay measurements. The abWSDGAT was shown to have the lowest activity at 0.02 U/mg and the mH WSDGAT had a higher activity at 0.07 U/mg. The purified mhWSDGAT was also used as a positive control (Figure 3.13).

3.2.5 FAS-B Activity Assay

3.2.5.1 Fatty Acid Extraction Method

Initial work had to consider which fatty acid extraction method to use. Two methods were available: the Dryer method (chloroform: methanol) or the ethyl-acetate method. The same

sample was divided and subjected to both methods, in duplicate. It was shown that the Dryer method was less noisy and more specific for the extraction of lipids and fatty acids. The ethyl-acetate method produced multiple peaks and other un-wanted products in the GC-chromatogram (Figure 3.14).

3.2.5.2 Detection of Fatty Acids from FAS

The production of palmitic acid (C_{16}) was tested in *S.cerevisiae* in YPD cultures grown for 72 hours. No palmitic acid was detected at all when cultures were grown in synthetic-dropout media, thus rich media was used in all characterizations. The cultures were subjected to fatty acid extraction through the Dryer method. Detection of the fatty acids was shown to not be reproducible and in very small quantities (Figure 3.15).

3.3 Discussions

3.3.1 Cloning

The *mhWSDGAT* and *abWSDGAT* genes were successfully cloned from the *M. hydrocarbonoclasticus* DSMZ 8798 and *A. baylyi* DSMZ 14959, respectively. However, after we cloned these genes, Shi *et al.* discovered a second WS enzyme from *M. hydrocarbonoclasticus* which had significant activity. Thus the WS2 gene was codon optimized and synthesized by DNA 2.0.

3.3.2 WSDGAT Activity

The assay to detect the WSDGAT activity was based the liberation of the sulfhydryl-CoA from the fatty alcohol and fatty acyl-CoA esterification. The reagent DTNB was used to react with the

released sulfhydryl-CoA. However, DTNB will react with any sulfhydryl group, thus this assay will have a very large background noise, because there are many different sulfhydryl groups in the cell crude lysate that could react with the DTNB. It was interesting to find that the order of substrates added was shown to have significant effects on the overall result. The enzyme should be added last. This discovery was also recently published by Barney and coworkers (14). In this work, Barney suggests that this could indicate a mechanism of allosteric control of protein activity based on substrate levels within the cell.

Though the expression studies of the WSDGAT in *S.cerevisiae* did not yield positive results, the activity assay proved that the gene was functionally expressed. The activity may be low, but it is active. This functional expression was also recently demonstrated by Shi *et al.* (15). In this work, the WS2 and the abWSDGAT were functionally expressed in *S.cerevisiae* for biodiesel production. It was found that the WS2 produced the highest biodiesel titer. The mhWSDGAT was not tested in the study by Shi. The future work of the WSDGAT activity will include testing all three enzymes, WS2, mh-, ab-WSDGAT.

3.3.3 FAS-B Activity

The irreproducibility of the palmitic acid production was initially troublesome. Having reproducible results is very important and had become a significant part of this project. However, after careful analysis and discussions with the new post-doctorate in the laboratory whom has had greater experience with fatty acids, this irreproducibility is now accepted as standard operating procedure when working with fatty acids. This type of system is inherently error-prone and making it more reproducible is futile. Therefore, future work will need to focus

on methods which over-produce fatty acids enough to make the error-margin insignificant compared to the amount of fatty acids being produced.

Not being able to detect excess C_{16} was troublesome. There are many reasons why this could have occurred. It is possible that the incubation time was too long. The current incubation time was at 70 hours, however it is possible that at this time, all of the nutrients had been utilized in the media and the cells had reverted to using the C_{16} energy storage, thus degrading the fatty acids. It is also possible that the excess fatty acids were degraded by the endogenous system due to the toxicity of excess fatty acids.

3.4 Conclusions and Future Work

Future work for this product will combine the FAS-B/PPT1 and WSDGAT modules to look for FAEE production. We are choosing to try this new strategy because it is expected that any excess fatty acids produced by the heterologously expressed FAS are being degraded. Therefore, it is hypothesized that incorporating a driving force for the conversion of fatty acids to FAEE, it is more likely that the excess fatty acids can be quantified and will not be degraded. The FAEE is not endogenously produced by the cell, so small quantities of the FAEE being produced will not be lost in the error-margin of the fatty acid production. Additionally, the FAEE cannot be utilized by the cell and will not be degraded. This new strategy was also partially instigated by the recent discovery of functional WSDGAT expression in *S.cerevisiae* (15). This project successfully expressed the WSDGATs in *S.cerevisiae* and detected FAEEs. Therefore, this strategy is a better platform: detect the FAEEs and optimize production. We must also keep in mind the error-margin that is inherent with working in fatty acids. Therefore, future work will

need to focus on methods which over-produce fatty acids enough to make the error-margin insignificant compared to the amount of fatty acids being produced. The design of experiments (DOE) for this new approach will involve identifying FAEEs from the single expression of the WSDGATs in *S.cerevisiae*, as was previously shown by Shi *et al.* (15). The DOE will also look at the combined expression of the FAS-B+WSDGAT to identify if the FAEE production is increased with the FAS expression.

At this time, the WS2 was not tested for any activity; future work will include testing the WS2 for activity towards ethanol, using the liberated Co-A mechanism for quantification. We also expect to repeat the crude enzyme lysate testing of the mh- and ab- WSDGAT for a complete analysis of all three enzymes comparatively. The WS2 is also being included in the FAEE testing.

3.5 Materials and Methods

3.5.1 Cloning from cDNA

The bacteria *M. hydrocarbonoclasticus* and *A. baylyi* were grown in Luria Broth (LB) media. The total RNA was isolated from fresh samples using the Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA) following manufacturer‟s instructions. Isolated RNA samples were cleaned using TURBO DNA-free Kit (AMBION INC, Austin, TX) and then reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis kit (Roche, Mannheim, Germany) with the oligo-dT primer following the manufacturer‟s instructions. Gene specific primers were used for the qPCR analysis from the cDNA (Table 3.1).

3.5.2 Yeast Plasmid Construction

The DNA Assembler method was used for all yeast plasmid construction (19). Yeast expression plasmids were constructed in a two module system: module 1 involved expression of FAS-B and PPT1 while module 2 involved expression of the WSDGATs. In module 1, the plasmids were both a high and low copy version of: pRS424-TPIp-FAS-TPIt-TEFp-PPT1-TEFt. In module 2, there were three plasmids, in high and low copy versions of: pRS425-PYKp-mhWSDGAT-PYKt, pRS425-PYKp-abWSDGAT-PYKt, pRS425-PYKp-WS2WSDGAT-PYKt (Figure 3.8). The FAS-B and PPT1 were amplified from the pGM44 plasmid (4) using gene specific primers from Table 3.1. The WS2 gene was synthesized by DNA 2.0, while the other two WSDGATs from *M. hydrocarbonoclasticus* and *A. baylyi* were cloned from the cDNA. A final yeast plasmid system is currently being constructed, combining both module 1 and module 2 into a single plasmid, shown in Figure 3.9 using the DNA Assembler method.

3.5.3 E.coli Plasmid Construction

The WSDGAT genes were cloned into a pET-28a vector for protein purification. The genes were amplified from the pRS416- WSDGAT plasmid for the respective WSDGAT. The primers for the construction are shown in Table 3.1. The primers included unique restriction digestion sites *NdeI* and *XhoI* for ligation into pET-28a. As mhWSDGAT has a *NdeI* site located within the gene, the restriction digest *NcoI* had to be used. This primer was very long to incorporate the histidine tag and was used to amplify the mhWSDGAT gene for ligation using *NcoI* and *XhoI*. Standard ligation with New England Biolabs (Ipswich, MA) ligase was used, following the manufacturer‟s protocol. The ligation reaction was transformed through electroporation into

DH5α *E.coli* cells from the Cell Media Facility (University of Illinois Urbana-Champaign, Urbana, IL).

3.5.4 SDS-PAGE

E.coli BL21 cells were used to test expression through SDS-PAGE gels. The cells were inoculated to an OD of 0.2 and grown for 2 hours at 37 °C. The cells were then induced with 0.7mM IPTG and grown for 6 hours at 30 °C. A sample was taken, washed with sterile water, then concentrated to either 10x or 100x for a final resuspension volume of 100 µL of 100 mg/mL lysozyme (Sigma Aldrich, St. Louis MO) in 100 mM potassium phosphate buffer pH 7 and then frozen at -80°C. The sample was then subjected to two rounds of freeze-thaw cycles before being pelleted. Standard SDS-PAGE gel protocol was followed.

3.5.5 Western Blotting

S.cerevisiae cells were inoculated at 0.2 OD and grown in 50 mL rich media in baffled flasks overnight at 30 °C. Cells were washed twice with water and concentrated either 10x or 100x with Yeast Breaking Buffer (MP Biomedical, Solon, OH). 1 mL of the culture was aliquoted out to Yeast Fast Prep Matrix C columns (MP Biomedical). The cells were lysed using the Fast Prep, with 40 second intervals of shaking at a setting shaking speed of 6.0 M/s. This was repeated seven times, with five minutes on ice in between each interval. The sample was then spun down to separate the lysate soluble and insoluble fractions. The protein concentration was determined via the Bradford Assay (Thermo Fisher Scientific, Rockford IL). A standard Western blot protocol was followed.

3.5.6 Protein Purification

Proteins with a histidine-tag were purified via a Talon Co^{2+} resin (Clontech). *E.coli* BL21 cells were used to express the protein for purification. To express the cells, 600 mL of LB media + 100 mg/ml kanamycin was used and an overnight culture was used to inoculate to an OD of 0.2. The culture was allowed to grow to an OD of 0.8 at 37 °C and 250 rpm, before being induced with 0.7mM IPTG. The temperature was then reduced to 30 °C and incubated for 6 hours. The cells were then harvested washed twice with sterile water, pelleted, and stored at -80 °C until use. The thawed pellets were then lysed using an Avestine homogenizer (Ottawa Canada) in 100 mg/mL lysozyme in 100 mM potassium phosphate buffer pH7. The sample was then purified through standard manufacturer"s protein purification protocols for a Talon Co^{2+} resin column. The purified protein was stored at -80 °C until use. The enzyme concentration was tested by the Bradford Assay following the manufacturer"s protocol.

3.5.7 WSDGAT Activity Assay

The WSDGAT activity was tested using purified mhWSDGAT and also crude lysate of ab- and mh- WSDGAT in *S.cerevisiae.* For the crude lysate, the cells were inoculated at 0.2 OD and grown in 50 mL of the appropriate nutrient drop-out media in baffled flasks overnight at 30 °C. Cells were washed twice with water and concentrated 10x with Yeast Breaking Buffer (MP Biomedical, Solon, OH). 1 mL of the culture was aliquoted out to Yeast Fast Prep Matrix C columns (MP Biomedical). The cells were lysed using the Fast Prep, with 40 second intervals of shaking at a speed of 6.0 M/s. This was repeated seven times, with five minutes on ice in between each interval. The sample was then spun down to separate the lysate soluble and insoluble fractions. The protein concentration was determined via the Bradford Assay following

manufacturer‟s protocol. Ellman‟s reagent was used to test the sulfhydryl groups (20). The following conditions were used for the assay buffer: 100 mM potassium phosphate buffer pH 8, 1mM EDTA with 0.1% Tergitol NP-11 detergent, 100 mM $MgCl_2$, 10 mM DNTB, 2500 μM palmitoyl-CoA, and 2500 μM ethanol. 1 mL was then transferred to UV cuvettes and blanked on a Cary 300 UV-Visible spectrophotometer (Agilent Technologies, Santa Clara, CA). 10 μL of the crude lysate was added and the absorbance was monitored at 412 nm.

3.5.8 Fatty Acid Extraction

3.5.8.1 Ethyl-Acetate

The strains to be tested were grown in 5 mL YPD media overnight and then transferred to a 15-mL glass test tube. 50 μM C_{19} dissolved in ethanol was added as an internal standard. 5 mL of ethyl acetate was added using a glass pipette and 500 μL of 6N HCl. The sample was then vortexed for 10 seconds and agitated at 30 °C at 250 rpm for 30 minutes. The sample was then allowed to phase separate and the top phase was aspirated off and transferred to a glass scintillation vial. This organic phase was allowed to dry overnight in the fume hood. Then 500 μL of a 2:1 methanol: toluene solution was added using a glass pipette. 100 μL of 2N TMS-diazomethane was added. The mixture was shaken to mix and then allowed to derivatize for 1 hour at 37°C. The remaining sample was then transferred to a glass GC tube for analysis.

3.5.8.2 Dryer Method

The strains to be tested were grown in 5 mL YPD media overnight and transferred to a 15-mL glass test tube. 50 μM C_{19} dissolved in ethanol was added as an internal standard. 5 mL of a 2:1 solution of chloroform: methanol was added with a glass pipette and vortexed for 10 seconds and

agitated at 30 °C at 250 rpm for 30 minutes. The sample was then allowed to phase separate. The top phase was aspirated off and the bottom layer was transferred to the glass scintillation vial. The top layer was recovered for further extraction as needed. The bottom layer which had been transferred to the scintillation vial was then allowed to dry overnight in the fume hood. Then 500 µL of a 2:1 methanol: toluene solution was added using a glass pipette. 100 µL of 2N TMS-diazomethane was added. The mixture was shaken to mix and then allowed to derivatize for 1 hour at 37 °C. The remaining sample was then transferred to a glass GC tube for analysis.

3.5.9 GC-MS Analysis

The fatty acids were separated and quantified through a GC-MS-QP 2010 Plus (Shimadzu, Kyoto Japan). The separation was performed by a DB-WAX GC column with the following dimensions: 30mm x 0.25 mm internal diameter, 0.25 µm film thickness, from Agilent Technologies (Santa Clara, CA). A 1 µL portion was injected via split less injection at 250 °C at an initial pressure of 98 kPA and a total flow of 14.1 mL/min helium carrier gas. The chromatograph separation initially occurred at a temperature of 140 °C for five minutes and then increased at a rate of 5 °C per minute to 250 °C and then held for three minutes. The mass transfer line and ion source were at 250 and 200 °C respectively. The fatty acids were detected with an electron ionization method in scan mode from 30 to 150 *m/z*. The identification was achieved by comparison of retentions times and mass spectrum.

3.5.10 qPCR Analysis

Samples of cultures grown to mid-late exponential phase were used to compare relative mRNA expression levels of enzymes via quantitative PCR (qPCR). The total RNA was isolated from

fresh samples using the Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA) following manufacturer‟s instructions. Isolated RNA samples were cleaned using TURBO DNA-free Kit (AMBION INC, Austin, TX) and then reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis kit (Roche, Mannheim, Germany) with the oligo-dT primer following the manufacturer‟s instructions. The qPCR was performed with LightCycler 480 SYBR Green Master reagents (Roche) using the Roche Light Cycler® 480 System (Roche, Indianapolis, IN) using gene specific primers (Table 3.1). The relative abundance of the mRNA levels for the target genes was normalized to the asparagine-linked glycosylation 9 gene (*alg9*) expression to determine expression levels.

3.6 References

1. Kalscheuer, R., Stolting, T. and Steinbuchel, A. (2006) Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology*, **152**, 2529-2536.
2. Gago, G., Diacovich, L., Arabolaza, A., Tsai, S.C. and Gramajo, H. (2011) Fatty acid biosynthesis in *actinomycetes*. *FEMS Microbiol. Rev.*, **35**, 475-497.
3. Schweizer, E. and Hofmann, J. (2004) Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol Mol Biol Rev*, **68**, 501-517.
4. Stuible, H.P., Meier, S. and Schweizer, E. (1997) Identification, isolation and biochemical characterization of a phosphopantetheine:protein transferase that activates the two type-I fatty acid synthases of *Brevibacterium ammoniagenes*. *Eur. J. Biochem.*, **248**, 481-487.
5. Stuible, H.P., Meurer, G. and Schweizer, E. (1997) Heterologous expression and biochemical characterization of two functionally different type I fatty acid synthases from *Brevibacterium ammoniagenes*. *Eur. J. Biochem.*, **247**, 268-273.
6. Zha, W., Shao, Z., Frost, J.W. and Zhao, H. (2004) Rational pathway engineering of type I fatty acid synthase allows the biosynthesis of triacetic acid lactone from D-glucose *in vivo*. *J. Am. Chem. Soc.*, **126**, 4534-4535.
7. Dellomonaco, C., Clomburg, J.M., Miller, E.N. and Gonzalez, R. (2011) Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. *Nature*, **476**, 355-359.
8. Lennen, R.M., Braden, D.J., West, R.A., Dumesic, J.A. and Pfleger, B.F. (2010) A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in *Escherichia coli* and catalytic conversion to alkanes. *Biotechnol. Bioeng.*, **106**, 193-202.
9. Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S.B. and Keasling, J.D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, **463**, 559-562.
10. Kalscheuer, R., Luftmann, H. and Steinbuchel, A. (2004) Synthesis of novel lipids in *Saccharomyces cerevisiae* by heterologous expression of an unspecific bacterial acyltransferase. *Appl. Environ. Microbiol.*, **70**, 7119-7125.
11. Stoveken, T., Kalscheuer, R., Malkus, U., Reichelt, R. and Steinbuchel, A. (2005) The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from *Acinetobacter* sp. strain ADP1: characterization of a novel type of acyltransferase. *J. Bacteriol.*, **187**, 1369-1376.
12. Alvarez, A.F., Alvarez, H.M., Kalscheuer, R., Waltermann, M. and Steinbuchel, A. (2008) Cloning and characterization of a gene involved in triacylglycerol biosynthesis and identification of additional homologous genes in the oleaginous bacterium *Rhodococcus opacus* PD630. *Microbiology*, **154**, 2327-2335.
13. Holtzapple, E. and Schmidt-Dannert, C. (2007) Biosynthesis of isoprenoid wax ester in *Marinobacter hydrocarbonoclasticus* DSM 8798: identification and characterization of isoprenoid coenzyme A synthetase and wax ester synthases. *J. Bacteriol.*, **189**, 3804-3812.
14. Barney, B.M., Wahlen, B.D., Garner, E., Wei, J. and Seefeldt, L.C. (2012) Differences in substrate specificities of five bacterial wax ester synthases. *Appl. Environ. Microbiol.*, **78**, 5734-5745.

15. Shi, S., Valle-Rodriguez, J.O., Khoomrung, S., Siewers, V. and Nielsen, J. (2012) Functional expression and characterization of five wax ester synthases in *Saccharomyces cerevisiae* and their utility for biodiesel production. *Biotechnol. Biofuels*, **5**, 7.
16. Barney, B.M., Mann, R.L. and Ohlert, J.M. (2012) Identification of a residue affecting fatty alcohol selectivity in wax ester synthase. *Appl. Environ. Microbiol.*
17. Kalscheuer, R. and Steinbuchel, A. (2003) A novel bifunctional wax ester synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in *Acinetobacter calcoaceticus* ADP1. *J. Biol. Chem.*, **278**, 8075-8082.
18. Willis, R.M., Wahlen, B.D., Seefeldt, L.C. and Barney, B.M. (2011) Characterization of a fatty acyl-CoA reductase from *Marinobacter aquaeolei* VT8: a bacterial enzyme catalyzing the reduction of fatty acyl-CoA to fatty alcohol. *Biochemistry*, **50**, 10550-10558.
19. Shao, Z. and Zhao, H. (2009) DNA assembler, an in *vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.*, **37**, e16.
20. Ellman, G.L. (1959) Tissue sulfhydryl groups. *Arch. Biochem. Biophys.*, **82**, 70-77.

3.7 Tables

Table 3.1: The primers used in this study

FAS Primers			
TPI F	5'-	GACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGATATATCTAGGAACCCATCA	-3'
TPI R	5'-	AAACACATACATAAACTAAAAATGACTATTGGCATCTCTAACCACCGCCTGGTGAACCG	-3'
FAS F	5'-	TAAATCTATAACTACAAAAAACACATACATAAACTAAAAATGACTATTGGCATCTCTAAC	-3'
FAS R	5'-	CTACGGCAAGCCACCAGTGAGATTAATATAATTATATAAAAATATTATCTTCTTTTCTT	-3'
TPIt F	5'-	CAAGGCGTGGTCGCAAGCGATCTACGGCAAGCCACCAGTGAGATTAATATAATTATATAA	-3'
TPIt R	5'-	AAATTTCAACTGTTATATAATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCAG	-3'
TEFP	5'-	GTTGAGAAGATGTTCTTATCCAAATTTCAACTGTTATATAATAGCTTCAAAATGTTTC	-3'
TEF t	5'-	CTAAGTTTTAATTACAAAATGCTCGACAACCGTGAAGCGATGACCGTGGGTGTGGACTTG	-3'
PPT1	5'-	GAAAGAAAGCATAGCAATCTAATCTAAGTTTTAATTACAAAATGCTCGACAACCGTGAAG	-3'
PPT1 R	5'-	CTGCGGTACCAGCGGTAAGGAGATTGATAAGACTTTTCTAGTTGCATATCTTTTATAT	-3'
TEF t	5'-	CGCCACCGCGCAGTGCCTGCTGCGGTACCAGCGGTAAGGAGATTGATAAGACTTTTC	-3'
TEF t R	5'-	CTTTGATCGGCGCTATACGCGTCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAG	-3'
WSDGAT Primers			
PGKp F	5'-	CAAATACTTTGATCGGCGCTATACGCACAGATATTATAACATCTGCACAATAGGCATTTG	-3'
PGKp R - to ab	5'-	CAAGGAAGTAATTATCTACTTTTTACAACAAATATAAAACAATGCGCCCATTACATCCG	-3'
ab F	5'-	CAAGGAAGTAATTATCTACTTTTTACAACAAATATAAAACAATGCATCATCATCATCATC	-3'
ab r	5'-	GTAATTGCAAAGCAGGAAGATATTAAAACAGCCAATTAAATTGAATTGAATTGAAATCG	-3'
PGK p R to mh	5'-	TACAACAAATATAAAACAATGACGCCCCTGAATCCCACTGACCAGCTCTTTCTCTGGC	-3'
PGKp Rto WS2	5'-	TACAACAAATATAAAACA ATGAAGAGATTAGGTACTCTAGACGCTAGTTGGCTTGCAG	-3'
mH F	5'-	CAAGGAAGTAATTATCTACTTTTTACAACAAATATAAAACAATGACGCCCCTGAATCCC	-3'
mH R	5'-	GAGCTCAACGCCGGTCTGTAAATTGAATTGAATTGAAATCGATAGATCAATTTTTTTC	-3'
WS2 F	5'-	GGAAGTAATTATCTACTTTTTACAACAAATATAAAACAATGAAGAGATTAGGTACTCTAG	-3'
WS2 R	5'-	GCGTGCCCGTACTAGAAAGTAAATTGAATTGAATTGAAATCGATAGATCAATTTTTTTC	-3'
PGKt F to ab	5'-	GTAATTGCAAAGCAGGAAGATATTAAAACAGCCAATTAAATTGAATTGAATTGAAATC	-3'
PGKt F to mh	5'-	CAGGGTCTGGCAGAGCTGGAGCTCAACGCCGGTCTGTAAATTGAATTGAATTGAAATC	-3'
PGKt F to WS2	5'-	CTAATTCTGCCACCAAAGAAGCGTGCCCGTACTAGAAAGTAAATTGAATTGAATTGAAATC	-3'
PGKt R	5'-	GTATAGTGTATTCTTCCTGCCATGGCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCG	-3'
WSDGAT pET 28 primers			-3'
mh F	5'-	GATATA**CCATGG**GCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGACGCCCCTGAATCCCACTGACC	-3'
mh R	5'-	GAGCTCAACGCCGGTCTGTAA**CTCGAG**CACCACCAC	-3'
ab F	5'-	CAGC**CATATG**ACGCCCCTGAATCCCACTG	-3'
ab R	5'-	GATATTAAAACAGCCAATTAA**CTCGAG**CACCACCACCACCACC	-3'
WS2 F	5'-	GGCAGC**CATATG**AAGAGATTAGGTACTC	-3'
WS2 R	5'-	CGTGCCCGTACTAGAAAGTAA**CTCGAG**CACCACC	-3'

3.8 Figures

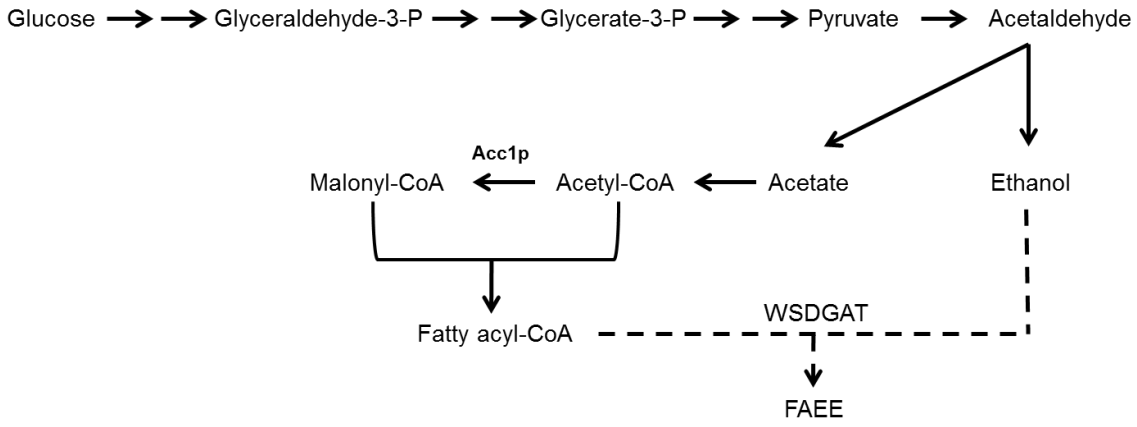


Figure 3.1: General scheme of microbial production of FAEEs (15).

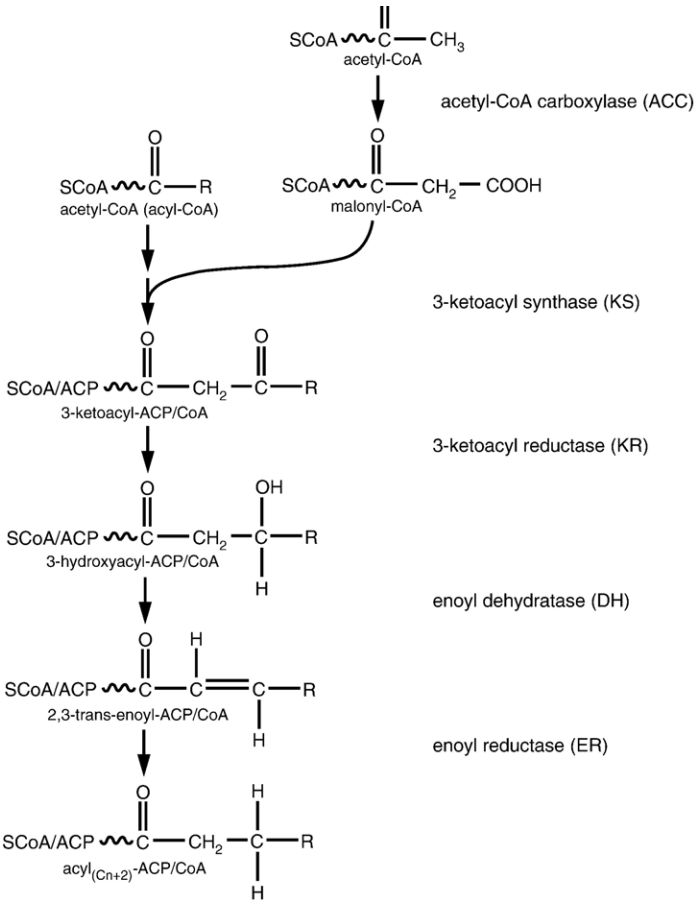


Figure 3.2: Mechanism of fatty acid biosynthesis (2).

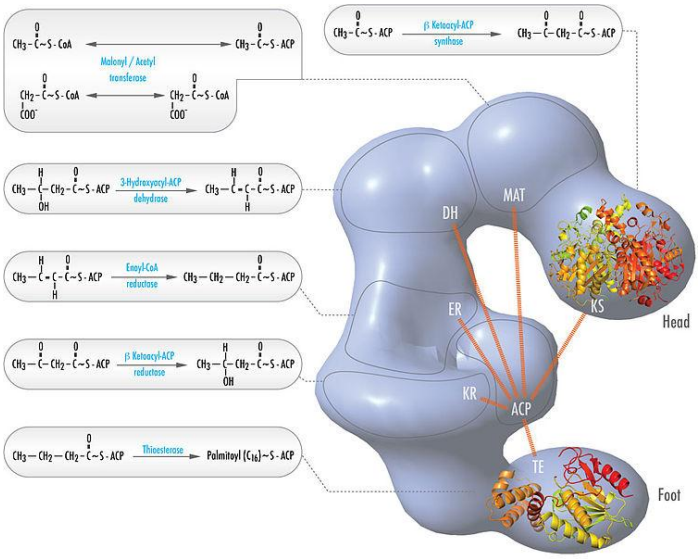


Figure 3.3: Schematic representation of the globular multi-functional enzyme complex FAS-B.

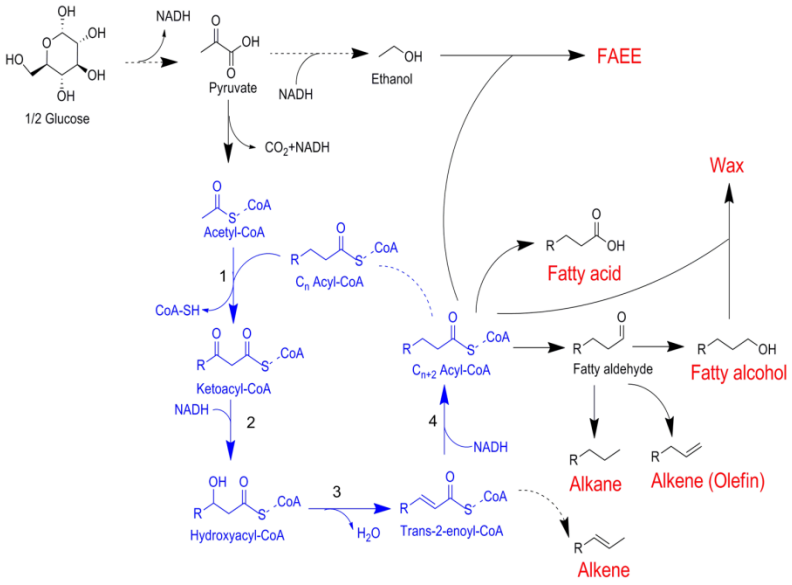


Figure 3.4: Reversal of the β-oxidation process. In this diagram, the enzymes are 1) 3-ketoacyl-CoA synthase; 2) 3-ketoacyl-CoA reductase; 3) 3-hydroxyacyl-CoA dehydratase 4) trans-2-enoyl-CoA reductase.

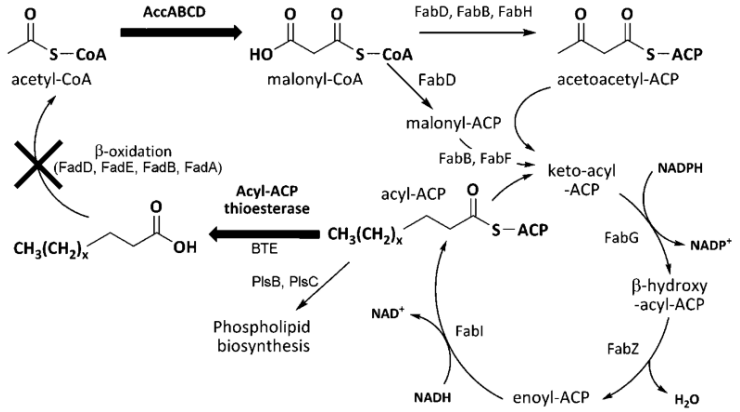


Figure 3.5: Metabolic engineering strategies used to overproduce fatty acids (8). Overexpression of the ACC allowed for the first committed step of the fatty acid pathway, pushing the flux towards fatty acids. Overexpression of the acyl-ACP thioesterase pulled the intermediates out of the cyclical fatty acid pathway. Also, knocking out *fadD* eliminated the breakdown of fatty acids in the β-oxidation pathway.

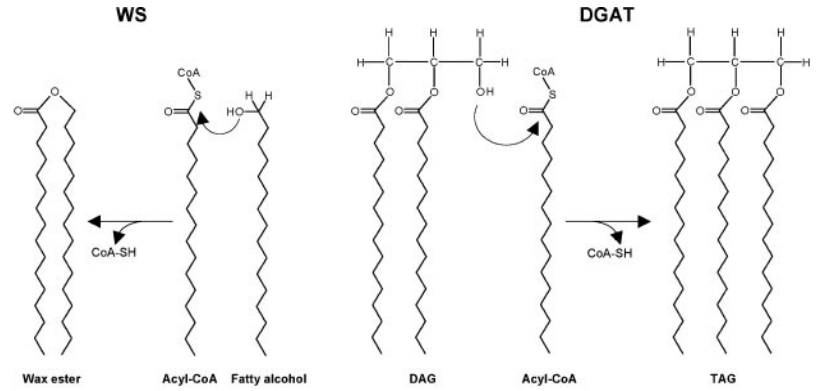


Figure 3.6: The WSDGAT is a bifunctional enzyme, with two reaction capabilities: wax ester synthase and AcylA:diacylglycerol transferase. We are interested in the wax synthase (WS) reaction (17).

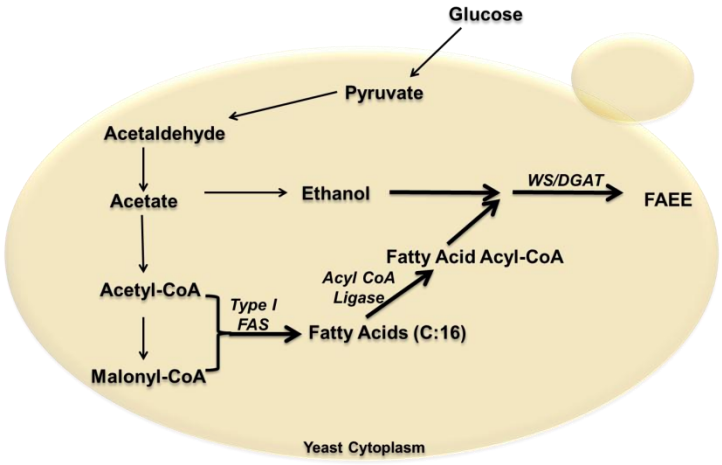


Figure 3.7: The general scheme for this research project. The heterologous expression of a bacterial Type I FAS is expected to overproduce the fatty acids because it will not be limited by endogenous regulation mechanisms. The overexpressed C_{16} will then be converted to FAEE by WSDGAT.

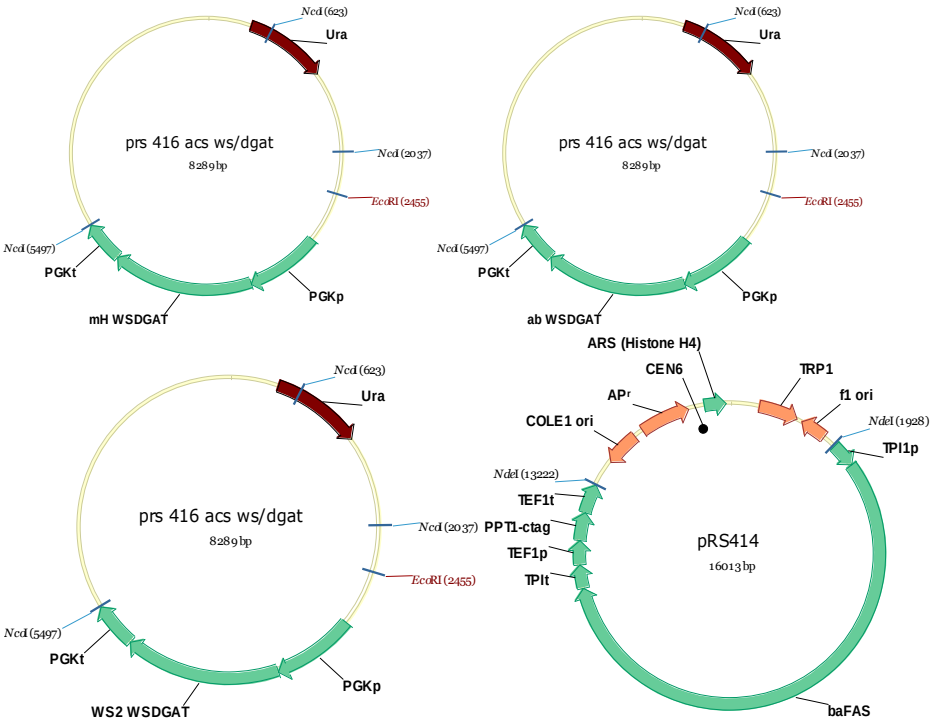


Figure 3.8: The initial plasmids created with individual enzymes WSDGAT and the FAS-B for activity analysis.

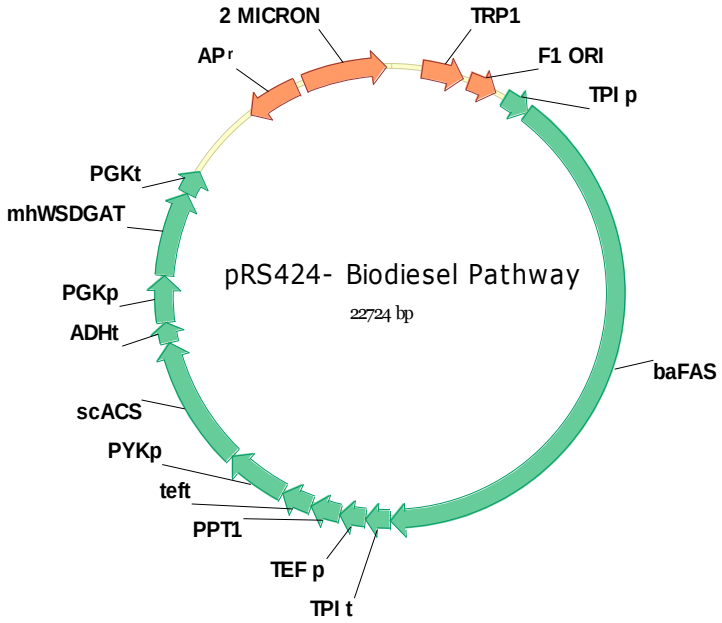


Figure 3.9: The final full biodiesel pathway which will consist of the FAS-B and the WSDGAT enzymes to produce FAEEs.

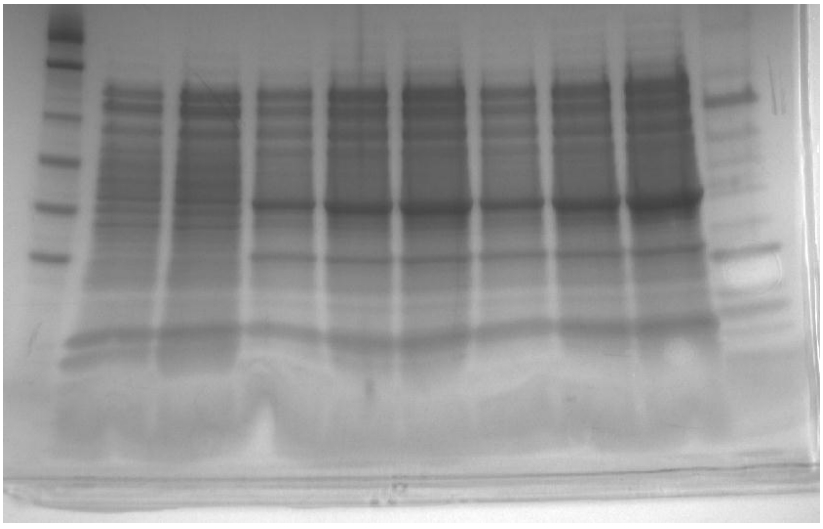


Figure 3.10: SDS-PAGE analysis of the mhWSDGAT and abWSDGAT enzymes expressed in *E.coli*. Lane 1: Ladder, Lane 2-3: Negative controls, Lane 4-6: mhWSDGAT Lane 7-9: abWSDGAT.

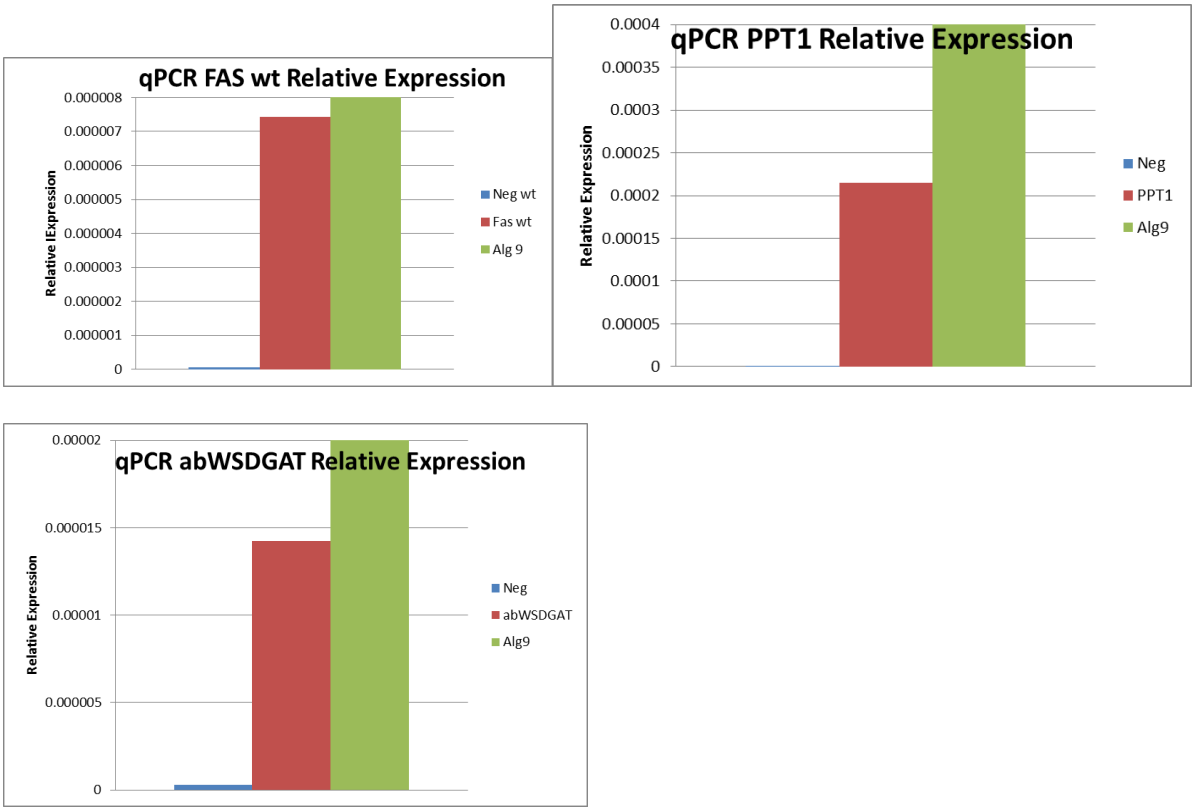


Figure 3.11: qPCR analysis of the FAS-B gene, the PPT1 gene, and the abWSDGAT gene.

Figure 3.12: Scheme of the Ellman"s reaction. DTNB reacts with the sulfuryl group to create a measureable yellow color. This is used in WSDGAT activity analysis to detect the release of the sulfuryl-CoA.

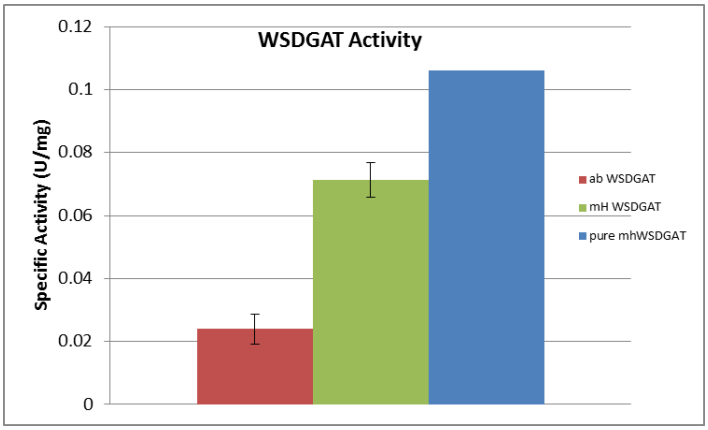


Figure 3.13: The data of the WSDGAT Ellman's reagent reaction. The negative was very high, but has been subtracted. The abWSDGAT and mhWSDGAT are both active.

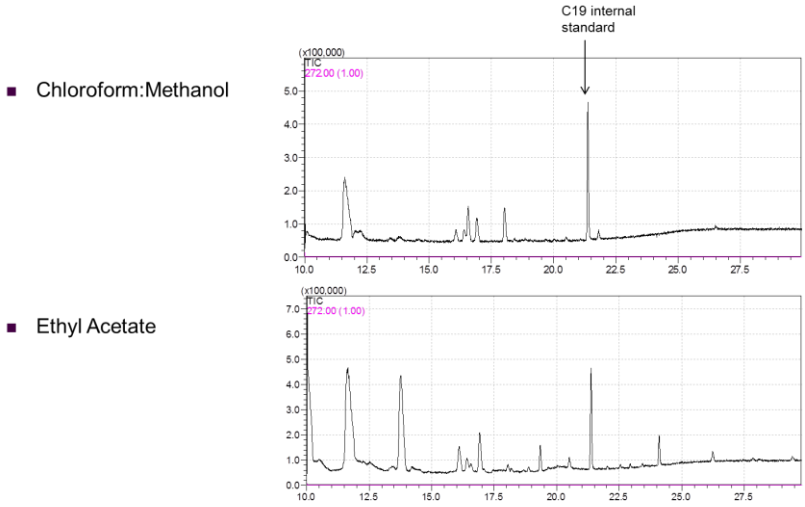


Figure 3.14: The GC chromatograms showing the two methods to extract fatty acids. The top method is the Dryer method (chloroform:methanol) and the second chromatogram is the ethyl acetate method. It is shown that the Dryer method is less noisy.

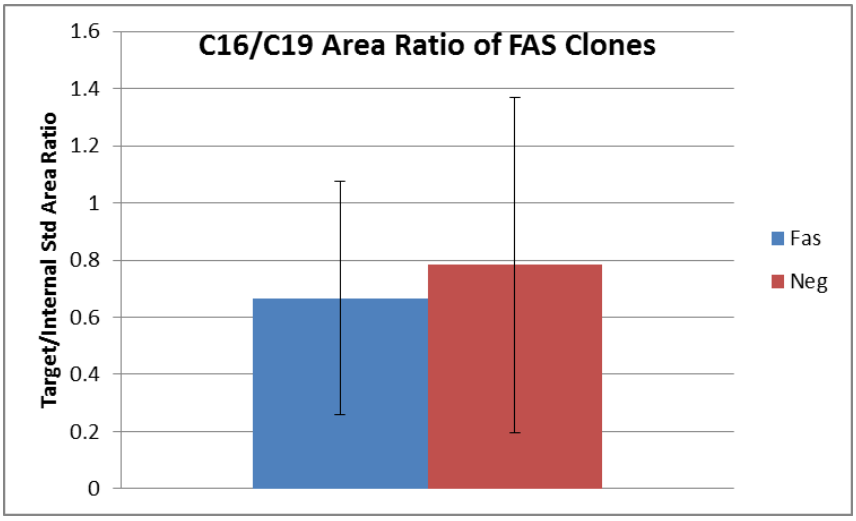


Figure 3.15: Detection of the C_{16} compound using GC-MS. The negative control is the strain harboring an empty plasmid. The FAS strain is expressing the fatty acid synthase FAS-B.